



[illegible]

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,281  
FILING DATE: 02-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/080,234  
FILING DATE: 03-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/826,622  
FILING DATE: 03-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 015280-295200US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..97  
OTHER INFORMATION: /product="OTHER"  
OTHER INFORMATION: /note="Gly at positions 1-97 may be  
OTHER INFORMATION: present or absent"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 105..201  
OTHER INFORMATION: /product="OTHER"  
OTHER INFORMATION: /note="Gly at positions 105-201 may be  
OTHER INFORMATION: present or absent"  
PS-09-054-281-22

```

Query Match      57.0%; Score 147; DB 4; Length 201;
Best Local Similarity 56.5%; Pred. No. 1.2e-06;
Matches 26; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 2 GGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSG 47
Db . | | | | | | | | | | | | | | | | | | | |
58 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 103

```

```

RESULT 7
US-09-478-948-6
; Sequence 6, Application US/09478948
; Patent No. 6465258
; GENERAL INFORMATION:
; APPLICANT: Shan, Bei
; APPLICANT: Okamoto, Arthur Y.
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: FXR Receptor-Mediated
; Modulation of Cholesterol
; TITLE OF INVENTION: Metabolism
; FILE REFERENCE: 018781-001310US
; CURRENT APPLICATION NUMBER: US/09/478,948
; CURRENT FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: US 60/115,249
; PRIOR FILING DATE: 1999-01-07

```

LOCATION: 6..40 "amino acid residues 6-40 may be



ADDRESSES: Nixon, Hargrave, Devans & Doyle  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York

```

;
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,254
; FILING DATE: 10-JUL-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-891-254-7

Query Match 48.3%; Score 124.5; DB 1; Length 344;
Best Local Similarity 40.6%; Pred. No. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps
QY 3 SGGSGGSGGSGGGG-----SGGSGGGGSGG-SGGSGGGSGGSGG 43
Db 141 GNGVGGANGAKAGGGGGLEALQETIEQLAQLGSGGAGGAGGGVGGAGGADGGSGAG 200
QY 44 GQSG 47
Db 201 GAGG 204

```

```

CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,270A
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,226
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1521
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-030-270A-7

Query Match 48.3%; Score 124.5; DB 2; Length 344;
Best Local Similarity 40.6%; Pred. No. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps
3

QY 3 GSGSGGQSGGSGSGGQ-----SGSGSGGQSGGQ-SGSGGQSGGSGGSGG 43
Db 141 GNGVGANGAGGAGGQGGGLAFALQETEQILALGGGAGAGAGGAGGAGGAGGAG 200
QY 44 GQSG 47
Db 201 GAGG 204

RESULT 15
US-08-984-207-7
; Sequence 7, Application US/08984207
; Patent No. 6235974
; GENERAL INFORMATION:
; APPLICANT: Oiu, Dewen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P. O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,207
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

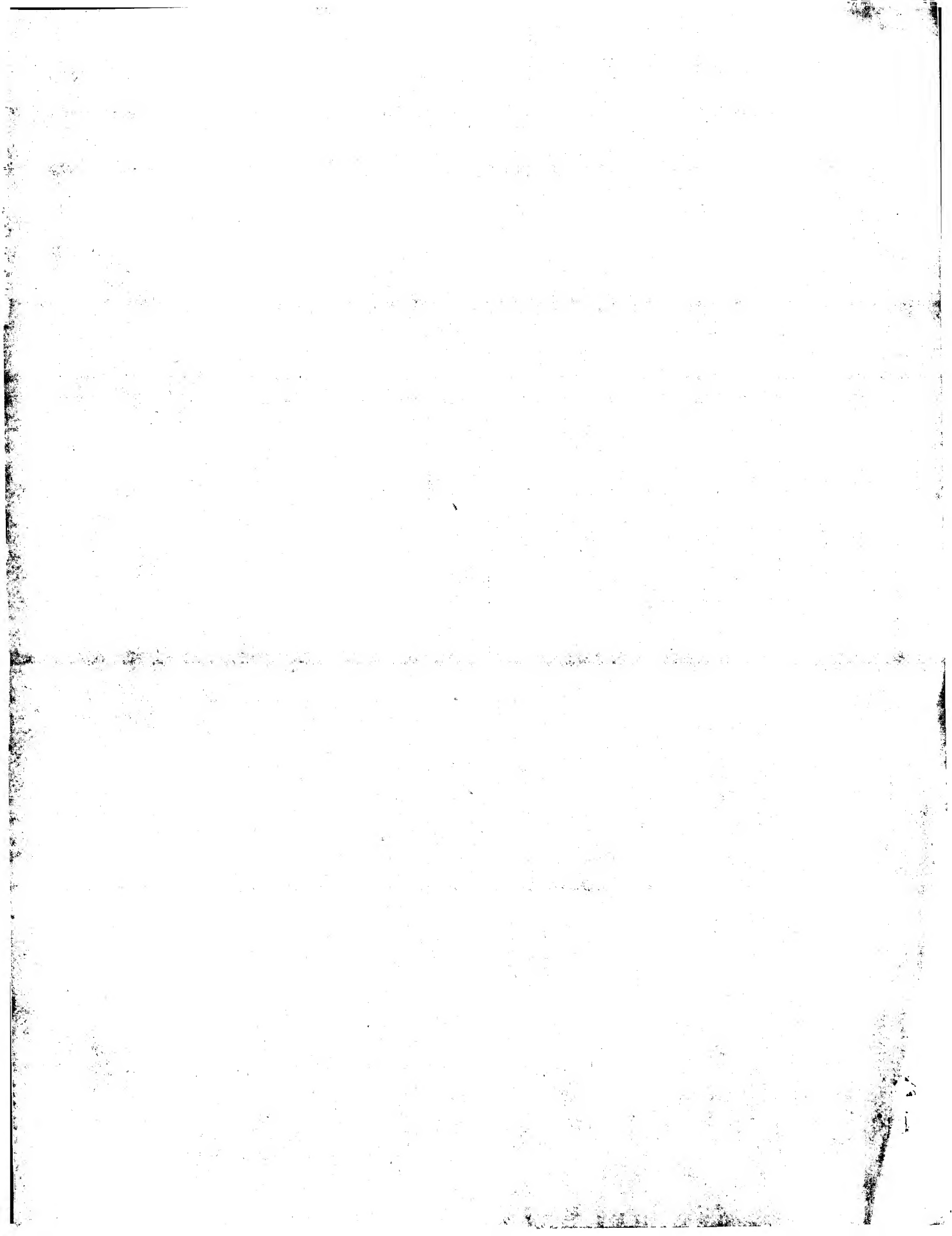
APPLICATION NUMBER: US 60/033,230  
FILING DATE: 05-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/1201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-984-207-7

Query Match 48.3%; Score 124.5; DB 4; Length 344;  
Best Local Similarity 40.6%; Pred. No. 0.00017;  
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;

QY 3 GSGSGGSGSGSGGQ-----SGSGSGSGSGG-SGSGSGSGSGSG 43  
Db 141 GNVGGANGAKGAGGGGLAEALQETIEQLAOLGGGGAGAGGGGAGGADGGSGAG 200

QY 44 GSGG 47  
Db 201 GAGG 204

Search completed: March 13, 2003, 15:06:20  
Job time : 3.37192 secs



[illegible]



(CNRS ) CENT NAT RECH SCI.

Baubet V, Le Mouellie H, Brulet P;  
WPI: 2002-139605/18.  
N-PSDB: ABA97889.

New fusion protein comprising a modified bioluminescent system with a fluorescent molecule covalently linked with a photoprotein, useful for monitoring calcium fluxes or for detecting electrical activity in a group of neural cells

Claim 6; Page 31; 6lpp; English.

The invention relates to a fusion protein (ABB08630-ABB08635) for energy transfer from aequorin to green fluorescent protein by Chemiluminescence Resonance Energy Transfer (CRET), which comprises a fluorescent molecule covalently linked with a photoprotein. The fusion protein comprises the formula: GFP LINKER - AEG

GFP = green fluorescent protein; AEG = aequorin; and LINKER = a polypeptide of 4-63 amino acids.

The fusion protein is useful as a bioluminescent Ca<sup>+</sup> reporter at the single cell level. The fusion protein, composition or bioluminescent system is useful for monitoring calcium fluxes in real time. This is particularly useful for understanding the development, the plasticity and the functioning of the central nervous system. The fusion protein comprising the bioluminescent system is useful for detecting electrical activity in a group of neural cells, for making it possible to complete the phenotype study of mutants, or for observing the calcium activity in a population of connected cells, for example in a neural network.

Query Match 100.0%; Score 258; DB 23; Length 477;  
Best Local Similarity 100.0%; Pred. No. 2.4e-16;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 SGGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 50  
|||||  
Db 239 SGGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 288

RESULT 4  
ABB08635  
ID ABB08635 standard; Protein; 906 AA.  
AC ABB08635;  
XX  
XX 03-MAY-2002 (first entry)  
XX  
XX GFP fusion protein Seg5A SEQ ID NO 6.  
XX  
XX  
XX GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;  
KW Chemiluminescence Resonance Energy Transfer; CRET;  
KW central nervous system; neural network.  
XX  
XX Synthetic.  
XX  
XX  
XX Key Location/Qualifiers  
PH Misc-difference 305  
FT /note= "Encoded by QGT"  
FT Misc-difference 623  
FT /note= "Encoded by CTQ"  
FT Misc-difference 868  
FT /note= "Encoded by GTQ"  
XX  
XX WO200192300-A2.  
XX  
XX  
XX PN  
XX  
XX PD  
XX 06-DEC-2001.  
XX  
XX 01-JUN-2001; 2001WO-EP07057.  
XX  
XX 01-JUN-2000; 2000US-208314P.

PR	09-JUN-2000; 2000US-210526P.
PR	14-DEC-2000; 2000US-255111P.
XX	
XX	
PA	(INSP ) INST PASTEUR.
PA	(CNRS ) CENT NAT RECH SCI.
XX	
XX	Baubet V, Le Mouellic H, Brulet P;
PI	WPI; 2002-139605/18.
XX	N-PSDE; ABA97890.
DR	
DR	
XX	
PT	New fusion protein comprising a modified bioluminescent system with a
PT	fluorescent molecule covalently linked with a photoprotein, useful for
PT	monitoring calcium fluxes or for detecting electrical activity in a
PT	group of neural cells
XX	
XX	
PS	Claim 7; Page 31-32; 61pp; English.
XX	
CC	The invention relates to a fusion protein (ABB08630-ABB08635) for energy
CC	transfer from aequorin to green fluorescent protein by Chemiluminescence
CC	Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
CC	covalently linked with a photoprotein. The fusion protein comprises the
CC	formula: GFP - LINKER - AEQ
CC	GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
CC	polypeptide of 4-63 amino acids.
CC	The fusion protein is useful as a bioluminescent Ca+ reporter at the
CC	single cell level. The fusion protein, composition or bioluminescent
CC	system is useful for monitoring calcium fluxes in real time. This is
CC	particularly useful for understanding the development, the plasticity and
CC	the functioning of the central nervous system. The fusion protein
CC	comprising the bioluminescent system is useful for detecting electrical
CC	activity in a group of neural cells, for making it possible to complete
CC	the phenotype study of mutants, or for observing the calcium activity in
CC	a population of connected cells, for example in a neural network.
XX	
XX	
SQ	Sequence 906 AA;
	Query Match 100.0%; Score 258; DB 23; Length 906;
	Best Local Similarity 100.0%; Pred. No. 4.1e-16;
	Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50
Db	668 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 717
RESULT 5	
ABB08633	
ID	ABB08633 standard; Protein: 468 AA.
XX	
AC	ABB08633;
XX	
DT	03-MAY-2002 (first entry)
XX	
DE	GFP fusion protein G4A SEQ ID NO 4.
XX	
KW	GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
KW	Chemiluminescence Resonance Energy Transfer; CRET;
KW	central nervous system; neural network.
XX	
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 401
FT	/note= "Encoded by QCA"
XX	
XX	WC200192300-A2.
PN	
XX	
PD	06-DEC-2001.
XX	
XX	01-JUN-2001; 2001WO-EP07057.
XX	
PR	01-JUN-2000; 2000US-208314P.

PR	09-JUN-2000; 2000US-210526P.
PR	14-DEC-2000; 2000US-255111P.
XX	
XX	
PA	(INSP ) INST PASTEUR.
PA	(CNRS ) CENT NAT RECH SCI.
XX	
XX	Baubet V, Le Mouellic H, Brulet P;
PI	WPI; 2002-139605/18.
XX	N-PSDE; ABA97890.
DR	
DR	
XX	
PT	New fusion protein comprising a modified bioluminescent system with a
PT	fluorescent molecule covalently linked with a photoprotein, useful for
PT	monitoring calcium fluxes or for detecting electrical activity in a
PT	group of neural cells
XX	
XX	
PS	Claim 7; Page 31-32; 61pp; English.
XX	
CC	The invention relates to a fusion protein (ABB08630-ABB08635) for energy
CC	transfer from aequorin to green fluorescent protein by Chemiluminescence
CC	Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
CC	covalently linked with a photoprotein. The fusion protein comprises the
CC	formula: GFP - LINKER - AEQ
CC	GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
CC	polypeptide of 4-63 amino acids.
CC	The fusion protein is useful as a bioluminescent Ca+ reporter at the
CC	single cell level. The fusion protein, composition or bioluminescent
CC	system is useful for monitoring calcium fluxes in real time. This is
CC	particularly useful for understanding the development, the plasticity and
CC	the functioning of the central nervous system. The fusion protein
CC	comprising the bioluminescent system is useful for detecting electrical
CC	activity in a group of neural cells, for making it possible to complete
CC	the phenotype study of mutants, or for observing the calcium activity in
CC	a population of connected cells, for example in a neural network.
XX	
XX	
SQ	Sequence 906 AA;
	Query Match 100.0%; Score 258; DB 23; Length 906;
	Best Local Similarity 100.0%; Pred. No. 4.1e-16;
	Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50
Db	668 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 717
RESULT 5	
ABB08633	
ID	ABB08633 standard; Protein: 468 AA.
XX	
AC	ABB08633;
XX	
DT	03-MAY-2002 (first entry)
XX	
DE	GFP fusion protein G4A SEQ ID NO 4.
XX	
KW	GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
KW	Chemiluminescence Resonance Energy Transfer; CRET;
KW	central nervous system; neural network.
XX	
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 401
FT	/note= "Encoded by QCA"
XX	
XX	WC200192300-A2.
PN	
XX	
PD	06-DEC-2001.
XX	
XX	01-JUN-2001; 2001WO-EP07057.
XX	
PR	01-JUN-2000; 2000US-208314P.

PR 09-JUN-2000; 2000US-210526P.  
 XX 14-DEC-2000; 2000US-255111P.  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CENT NAT RECH SCI.  
 XX Baubet V, Le Mouellic H, Brulet P;  
 PI WPI; 2002-139605/18.  
 XX New fusion protein comprising a modified bioluminescent system with a  
 PT fluorescent molecule covalently linked with a photoprotein, useful for  
 PT monitoring calcium fluxes or for detecting electrical activity in a  
 PT group of neural cells  
 XX Claim 5; Page 31; 61pp; English.  
 XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy  
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence  
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule  
 CC covalently linked with a photoprotein. The fusion protein comprises the  
 CC formula: GFP - LINKER - AEQ  
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a  
 CC polypeptide of 4-63 amino acids.  
 CC The fusion protein is useful as a bioluminescent Ca+ reporter at the  
 CC single cell level. The fusion protein, composition or bioluminescent  
 CC system is useful for monitoring calcium fluxes in real time. This is  
 CC particularly useful for understanding the development, the plasticity and  
 CC the functioning of the central nervous system. The fusion protein  
 CC comprising the bioluminescent system is useful for detecting electrical  
 CC activity in a group of neural cells, for making it possible to complete  
 CC the phenotype study of mutants, or for observing the calcium activity in  
 CC a population of connected cells, for example in a neural network.  
 XX Sequence 468 AA;  
 SQ Query Match 82.2%; Score 212; DB 23; Length 468;  
 Best Local Similarity 97.6%; Pred. No. 3.5e-12;  
 Matches 41; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 QSGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 50  
 :|||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 238 KSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 279  
 RESULT 6  
 ABB08756  
 ID ABB08756 standard; peptide; 41 AA.  
 XX ABB08756;  
 AC ABB08756;  
 XX 03-MAY-2002 (first entry)  
 DT GFP peptide linker SEQ ID NO 21.  
 XX GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;  
 KW Chemiluminescence Resonance Energy Transfer; CRET;  
 KW central nervous system; neural network.  
 XX Synthetic.  
 OS WO200192300-A2.  
 XX 06-DEC-2001.  
 PD 01-JUN-2001; 2001WO-EP07057.  
 XX 01-JUN-2000; 2000US-208314P.  
 PR 09-JUN-2000; 2000US-210526P.  
 PR 14-DEC-2000; 2000US-255111P.  
 XX (INSP ) INST PASTEUR.  
 PA

PA (CNRS ) CENT NAT RECH SCI.  
 XX Baubet V, Le Mouellic H, Brulet P;  
 XX WPI; 2002-139605/18.  
 XX New fusion protein comprising a modified bioluminescent system with a  
 PT fluorescent molecule covalently linked with a photoprotein, useful for  
 PT monitoring calcium fluxes or for detecting electrical activity in a  
 PT group of neural cells  
 XX Claim 23; Page 35; 61pp; English.  
 PS The invention relates to a fusion protein (ABB08630-ABB08635) for energy  
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence  
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule  
 CC covalently linked with a photoprotein. The fusion protein comprises the  
 CC formula: GFP - LINKER - AEQ  
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a  
 CC polypeptide of 4-63 amino acids.  
 CC The fusion protein is useful as a bioluminescent Ca+ reporter at the  
 CC single cell level. The fusion protein, composition or bioluminescent  
 CC system is useful for monitoring calcium fluxes in real time. This is  
 CC particularly useful for understanding the development, the plasticity and  
 CC the functioning of the central nervous system. The fusion protein  
 CC comprising the bioluminescent system is useful for detecting electrical  
 CC activity in a group of neural cells, for making it possible to complete  
 CC the phenotype study of mutants, or for observing the calcium activity in  
 CC a population of connected cells, for example in a neural network. The  
 CC present sequence is that of a GFP peptide linker sequence of the  
 CC invention.  
 XX Sequence 41 AA;  
 SQ Query Match 81.8%; Score 211; DB 23; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-13;  
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 SGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 50  
 :|||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 1 SGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 41  
 RESULT 7  
 AAW56163  
 ID AAW56163 standard; Protein; 738 AA.  
 XX AAW56163;  
 AC AAW56163;  
 XX 28-JUL-1998 (first entry)  
 DT New DNA sequence isolated from Pinctada fucata.  
 DE Pinctada fucata; protein; cosmetic.  
 XX Pinctada fucata.  
 OS JP10080285-A.  
 XX JP10080285-A.  
 XX 31-MAR-1998.  
 PD 28-MAY-1997; 97JP-0138461.  
 XX 15-JUL-1996; 96JP-0184459.  
 XX (MIKI-) MIKIMOTO SEIYAKU KK.  
 PA WPI; 1998-254410/23.  
 DR N-PSDB; AAW22683.  
 XX New cDNA and e.g. vector, host cell and polypeptide - used to  
 PT produce polypeptide in high yields, which is used in cosmetics  
 XX









[illegible]

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	147.5	57.2	532	10	US-09-833-790-428	Sequence 428, Appl	
2	147	57.0	201	10	US-09-848-990-22	Sequence 22, Appl	
3	147	57.0	201	10	US-09-760-364-14	Sequence 14, Appl	
4	147	57.0	357	10	US-09-864-761-35807	Sequence 35807, A	
5	146	56.6	200	9	US-10-160-354-4	Sequence 4, Appl	
6	146	56.6	200	9	US-09-990-940-21	Sequence 21, Appl	
7	146	56.6	200	9	US-10-026-021-8	Sequence 8, Appl	
8	146	56.6	200	9	US-10-161-165-3	Sequence 3, Appl	
9	146	56.6	200	9	US-10-160-563-3	Sequence 3, Appl	
10	146	56.6	200	9	US-10-071-838-15	Sequence 15, Appl	
11	146	56.6	200	9	US-10-094-417-35	Sequence 25, Appl	
12	146	56.6	200	10	US-09-798-584-18	Sequence 18, Appl	
13	146	56.6	200	10	US-09-967-624-19	Sequence 19, Appl	
14	146	56.6	200	10	US-09-998-667-18	Sequence 18, Appl	
15	144	55.8	40	9	US-10-081-400-1	Sequence 1, Appl	
16	139.5	54.1	618	10	US-09-925-500-1381	Sequence 1381, Ap	
17	138	53.5	484	9	US-09-820-843A-19	Sequence 19, Appl	
18	137	51.1	283	10	US-09-864-761-36720	Sequence 36720, A	
19	136.5	52.9	40	9	US-10-005-438-9	Sequence 9, Appl	



```

; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-09-990-940-21

Query Match          56.6%; Score 146; DB 9; Length 200;
Best Local Similarity 56.5%; Pred. No. 1.5e-05;
Matches 26; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      2 GSGSGGGGCGSGGCGSGGCGSGGCGSGGCGSGGCGSGGCGSQG 47
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
DB      103 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148

RESULT 7
US-10-026-021-8
; Sequence 8, Application US/10026021
; Publication No. US20030027756A1
; GENERAL INFORMATION:
; APPLICANT: Hitoshi, Yasumichi
; APPLICANT: Demo, Susan
; APPLICANT: Jenkins, Yonchu
; APPLICANT: Rigel Pharmaceuticals, Inc.
; TITLE OF INVENTION: SAK: Modulation of Cellular Proliferation for
; TITLE OF INVENTION: Treatment of Cancer
; FILE REFERENCE: 021044-001210US
; CURRENT APPLICATION NUMBER: US/10/026,021
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/309,632
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-10-026-021-8

Query Match          56.6%; Score 146; DB 9; Length 200;
Best Local Similarity 56.5%; Pred. No. 1.5e-05;
Matches 26; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      2 GSGSGGGGCGSGGCGSGGCGSGGCGSGGCGSGGCGSGGCGSQG 47
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
DB      103 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148

RESULT 8
US-10-161-165-3
; Sequence 3, Application US/10161165
; Publication No. US20030027763A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Mark
; APPLICANT: Holland, Sacha
; APPLICANT: Rossi, Alex
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: CD43: Modulators of Mast Cell Degranulation
; FILE REFERENCE: 021044-001010US
; CURRENT APPLICATION NUMBER: US/10/161,165

```

```

; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,801
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
US-10-161-165-3

Query Match          56.6%; Score 146; DB 9; Length 200;
Best Local Similarity 56.5%; Pred. No. 1.5e-05;
Matches 26; Conservative 0; Mismatches 20; Indels 0; Gaps

QY    2   GGSGGGGSGGSGGGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 47
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    103  GGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148

RESULT 9
US-10-160-663-3
; Sequence 3, Application US/10160663
; Publication No. US20030040001A1
; GENERAL INFORMATION:
; APPLICANT: Demo., Susan
; APPLICANT: Hitoshi Yasumichi
; APPLICANT: Pearsall, Denise
; APPLICANT: Rigel Pharmaceutical, Incorporated
; TITLE OF INVENTION: LETM1: Modulators of Cellular Proliferation
; FILE REFERENCE: 021044-000920US
; CURRENT APPLICATION NUMBER: US/10/160,663
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,817
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/347,970
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
US-10-160-663-3

Query Match          56.6%; Score 146; DB 9; Length 200;
Best Local Similarity 56.5%; Pred. No. 1.5e-05;
Matches 26; Conservative 0; Mismatches 20; Indels 0; Gaps

QY    2   GGSGGGGSGGSGGGGSGGSGGSGGSGGSGGSGGSGGSGGSG 47
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    103  GGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148

RESULT 10
US-10-071-838-15
; Sequence 15, Application US/10071838
```



Query Match	56.6%	Score 146;	DB 10;	Length 200;
Best Local Similarity	56.5%	Pred. No. 1.5e-05;		
Matches 26;	Conservative	0;	Mismatches 20;	Indels 0;
				Gaps 0;

[illegible]

```

:
: LENGTH: 200
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
:

```

```

:
: OTHER INFORMATION: Description of Artificial Sequence:Flexible linker
:
: FEATURE:
: NAME/KEY: MOD_RES
: LOCATION: (6_..700)
:

```

US-09-998-667-18

Query Match	56.6%	Score 146;	DB 10;	Length 200;
Best Local Similarity	56.5%;	Pred. No. 1.5e-05;		
Matches 26;	Conservative	0;	Mismatches 20;	Indels 0;
Gaps 0;				

RESULT 15  
US-10-081-400-1  
; Sequence 1, Application US/10081400  
; Patent No. US20020155998A1

```

: CURRENT APPLICATION NUMBER: US/10/081,400
:
: CURRENT FILING DATE: 2002-02-20
:
: PRIOR APPLICATION NUMBER: US 09/333,213
:
: PRIOR FILING DATE: 1999-06-15
:
: NUMBER OF SEQ ID NOS: 4
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 1
:
: LENGTH: 40
:
: TYPE: PRT
:
: ORGANISM: Artificial Sequence
:
: FEATURE:
:

```



Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	170	65.9	302	2	C84470	hypothetical prote	
2	168	65.1	592	2	E82759	endo-1,4-beta-gluc	
3	161	62.4	115	2	D81615	sericin MG-1 - gre	
4	159.5	61.8	207	2	T07381	glycine-rich prote	
5	154	59.7	641	1	Q0BE31	nuclear antigen EB	
6	153	59.3	1218	2	E84537	hypothetical prote	
7	148.5	57.6	171	2	H84709	probable glycine-r	
8	146	56.6	291	1	S31415	glycine-rich prote	
9	146	56.6	622	2	I37984	keratin 9, type I,	
10	145	56.2	1226	2	T24045	hypothetical prote	
11	144.5	56.0	221	2	T04592	glycine-rich cell	
12	144	55.8	183	2	PN0109	keratin-like prote	
13	144	55.8	569	1	KRMS1	keratin, 59K type	
14	143.5	55.6	208	2	T46896	merozoite surface	
15	143.5	55.6	593	1	KRHU0	keratin 10, type I	
16	143	55.4	465	1	S01820	glycine-rich cell	
17	143	55.4	526	1	KRBOVI	keratin, 54K type	
18	142.5	55.2	165	1	KR2G1	glycine-rich cell	
19	142.5	55.2	166	1	KRBO2B	keratin, 68K type	
20	142.5	55.2	570	2	S07330	keratin, epidermal	
21	142	55.0	481	2	A35628	loricrin - mouse	
22	142	55.0	581	1	KRMS2	keratin, type II c	
23	141	54.7	396	2	T49109	glycine-rich prote	
24	139	53.9	183	1	KNR2G2	glycine-rich cell	
25	138.5	53.7	271	2	S34666	glycine-rich prote	
26	138.5	53.7	434	1	Z3BP1K	coat protein A - p	
27	138	53.5	484	2	G70846	hypothetical glyci	
28	137.5	53.3	274	2	T25404	hypothetical prote	
29	137	53.1	320	2	T09555	fibrillarlin - Arab	



A:Residues: 1-622 <RES>  
 A:Cross-references: EMBL:X75015; NID:9453154; PIDN:CAA52924.1; PID:9453155  
 R:Langbein, L.; Heid, H.W.; Moll, I.; Franke, W.W.  
 Differentiation 55, 57-71, 1993  
 A:Title: Molecular characterization of the body site-specific human epidermal cytoker  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 17-May-2002  
 C:Accession: H84709  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10611797  
 A:Accession: H84709  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-171 <STO>  
 A:Cross-references: GB:AE002093; NID:g1946363; PIDN:AAB63081.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g30560  
 A:Map position: 2  
 C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8  
 Query Match 57.6%; Score 148.5; DB 2; Length 171;  
 Best Local Similarity 63.0%; Pred. No. 2e-05;  
 Matches 29; Conservative 1; Mismatches 13; Indels 3; Gaps 1;  
 QY 2 GSGSGG---QSGSGSGSGSGSGSGSGSGSGSGSGSGG 44  
 Db 5 GSGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 50  
 RESULT 8  
 S31415  
 glycine-rich protein GRP22 - rape  
 C:Species: Brassica napus (rape)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S31415  
 R:Bergeron, D.; Boivin, R.; Baszczynski, C.L.; Bellemare, G.  
 submitted to the EMBL Data Library, August 1992  
 A:Description: Characterization and expression of a gene family encoding glycine-rich pr  
 A:Reference number: S31415  
 A:Accession: S31415  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-291 <BER>  
 A:Cross-references: EMBL:Z15045; NID:g17820; PIDN:CAA78762.1; PID:g17821  
 C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8  
 Query Match 56.6%; Score 146; DB 1; Length 291;  
 Best Local Similarity 53.7%; Pred. No. 4.9e-05;  
 Matches 29; Conservative 2; Mismatches 15; Indels 8; Gaps 1;  
 QY 2 GSGSGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGG 47  
 Db 80 GGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 133  
 RESULT 9  
 I37984  
 keratin 9, type I, cytoskeletal - human  
 N:Alternate names: cytokeratin 9; scatter protein 60K chain, placental  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 29-Oct-1999  
 C:Accession: I37984; S40307; S77921; S41161; B35494; I37943  
 R:Reis, A.; Hennies, H.C.; Langbein, L.; Digweed, M.; Mischke, D.; Drechsler, M.; Schrod  
 Nature Genet. 6, 174-179, 1994  
 A:Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).  
 A:Reference number: I37984; MUID:94214498; PMID:7512862  
 A:Accession: I37984  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Status: preliminary; translated from GB/EMBL/DDBJ



keratin 10, type I, cytoskeletal - human  
N:Alternate names: cytokeratin 10  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 10-Dec-1999  
C:Accession: S02158; C38182; B38182; PC1102; S14666; S14669  
R:Rieger, M.; Franke, W.W.  
J. Mol. Biol. 204, 841-856, 1988  
A:Title: Identification of an orthologous mammalian cytotokeratin gene. High degree of int  
A:Reference number: S02158; MUID:89125611; PMID:2464696  
A:Accession: S02158  
A:Molecule type: DNA  
A:Residues: 1-593 <RIE>  
A:Cross-references: EMBL:X14487; NID:q28316; PIDN:CAA32649.1; PID:q28317  
A:Experimental source: clone lambda-KH10-5  
R:Korge, B.P.; Gan, S.Q.; McBride, O.W.; Mischke, D.; Steinert, P.M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 910-914, 1992  
A:Title: Extensive size polymorphism of the human keratin 10 chain resides in the C-term  
A:Reference number: A38182; MUID:92141228; PMID:1371013  
A:Accession: C38182  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 452-593 <KOR1>  
A:Cross-references: PIDN:AA21315.1; PID:g244509  
A:Note: sequence extracted from NCBI backbone (NCBI:P79427)  
A:Accession: B38182  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 452-463, P, 465-507, Y, 523-593 <KOR2>  
A:Cross-references: PIDN:AA21314.1; PID:g244508  
A:Note: sequence extracted from NCBI backbone (NCBI:P79431)  
R:Kachenko, A.V.; Buchman, V.L.; Bliskovsky, V.V.; Shvets, Y.P.; Kisselev, L.L.  
Gene 116, 245-251, 1992  
A:Title: Exons I and VII of the gene (Ker10) encoding human keratin 10 undergo structural  
A:Reference number: PC1102; MUID:92339897; PMID:1378806  
A:Accession: PC1102  
A:Molecule type: mRNA  
A:Residues: 'G', 198-407, 'Q', 409-450, 'G', 452-485, 491-524, 534-593 <TKA>  
A:Cross-references: GB:M77663; NID:g186628; PIDN:AAA59199.1; PID:g186629  
A:Experimental source: embryonic skin, clone HK51  
R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.  
Mol. Biol. Rep. 12, 277-283, 1987  
A:Title: Sequence of a cDNA encoding human keratin No 10 selected according to structural  
A:Reference number: S14666; MUID:88122104; PMID:2448602  
A:Accession: S14666  
A:Molecule type: mRNA  
A:Residues: 130-278, 'YV', 281-311, 'I', 313-339, 'V', 341-373, 'R', 375-407, 'Q', 409-459, 'RS', 46  
56-579, 'P', 581-593 <DAR1>  
A:Cross-references: EMBL:M19156; NID:g186769  
A:Note: the sequence from Fig. 3 is inconsistent with the nucleotide sequence from Fig.  
R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.  
submitted to the EMBL Data Library, May 1988  
A:Reference number: S14667  
A:Accession: S14667  
A:Molecule type: mRNA  
A:Residues: 130-278, 'YV', 281-311, 'I', 313-339, 'V', 341-373, 'R', 375-407, 'Q', 409-459, 'RS', 46  
56-593 <DAR2>  
A:Cross-references: EMBL:M19156; NID:g186769; PIDN:AAA59468.1; PID:g307086  
A:Note: the translated sequence in GenBank entry HUMKRT10A, release 111.0, differs from  
C:Genetics:  
A:Gene: GDB:KRT10; KPP  
A:Cross-references: GDB:118828; OMIM:148080  
A:Map position: 17q12-17q21  
A:Introns: 209/3; 237/2; 289/3; 343/3; 385/3; 458/2; 592/3  
A:Note: this gene encodes variants with considerable length polymorphism  
C:Note: mutations in this gene can cause epidermolytic hyperkeratosis and keratosis palm  
C:Complex: heterotrimer of two type I and two type II proteins, usually keratin I (see  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; heterotrimer; intermediate filament; polymorphism





Result No.	Query			ID	Description
	Score	Match	Length		
1	161.5	62.6	450	1	SWPL_ENCCU
2	161	62.4	115	1	SER1_GALME
3	154	59.7	641	1	EBN1_EBV
4	147.5	57.2	532	1	KIC2_HUMAN
5	146	56.6	622	1	KIC1_HUMAN
6	144	55.8	569	1	KICJ_MOUSE
7	143.5	55.6	593	1	KICJ_HUMAN
8	143	55.4	465	1	GRP2_PHAVU
9	143	55.4	526	1	KICJ_BOVIN
10	143	55.4	677	1	SPB7_DICDI
11	142.5	55.2	165	1	GRP1_BOVIN
12	142.5	55.2	166	1	K2C5_BOVIN
13	142	55.0	481	1	LORI_MOUSE
14	141	54.7	627	1	K2C1_MOUSE
15	139	53.9	183	1	GRP2_ORYSA
16	138.5	53.7	434	1	COAA_BPKE
17	138	53.5	1250	1	TP3A_DROME
18	137	53.1	384	1	GRP1_PETHY
19	136	52.7	333	1	SLX3_MOUSE
20	135	52.3	214	1	GRD_NICYS
21	135	52.3	1901	1	Y208_MYCTU
22	133.5	51.7	316	1	LORI_HUMAN
23	133	51.6	252	1	GRP1_PHAVU
24	132	51.2	386	1	RB87_DROME
25	131.5	51.0	463	1	YAG8_MYCTU
26	131.5	51.0	491	1	YK98_MYCTU
27	131	50.8	1380	1	DX99_CAEEL
28	131	50.8	3178	1	YSB9_CAEEL
29	130.5	50.6	434	1	COAA_BP122
30	130.5	50.6	643	1	K2C1_HUMAN
31	130	50.4	440	1	FXGA_CHICK
32	130	50.4	763	1	GLH1_CAEEL
33	130	50.4	1726	1	MSP1_PLAFC

[illegible]

Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
TISSUE=Foot sole tissue;  
MEDLINE=94131202; PubMed=7507869;  
Langbein L., Held H.W., Moil I., Franke W.W.;  
RT "Molecular characterization of the body site-specific human epidermal  
cytokeratin 9: cDNA cloning, amino acid sequence, and tissue  
specificity of gene expression.";  
J. Invest. Dermatol. 103:474-477(1994).  
[2]  
SEQUENCE OF 449-465.  
MEDLINE=90267446; PubMed=2140676;  
Rosen E.M., Meromsky L., Romero R., Setter E., Goldberg I.;  
RT "Human placenta contains an epithelial scatter protein.";  
Biochem. Biophys. Res. Commun. 168:1082-1088(1990).  
[3]  
VARIANTS EPK VAL-156 AND PRO-171.  
MEDLINE=94274199; PubMed=7516304;  
Hennies H.-C., Zehender D., Kunze J., Kuester W., Reis A.;  
RT "Keratin 9 gene mutational heterogeneity in patients with  
epidermolytic palmoplantar keratoderma.";  
Hum. Genet. 93:649-654(1994).  
[4]  
VARIANT EPK SER-160.  
MEDLINE=95015968; PubMed=7523529;  
Bonifas J.M., Matsumura K., Chen M.A., Berth-Jones J.,  
RA Hutchinson P.E., Zlotcower M., Fritsch P.O., Epstein E.H. Jr.;  
RT "Mutations of keratin 9 in two families with palmoplantar  
epidermolytic hyperkeratosis.";  
J. Invest. Dermatol. 103:474-477(1994).  
[5]  
VARIANT EPK TYR-160.  
MEDLINE=94184355; PubMed=7511021;  
Torcador D., Blanchet-Bardon C., Serova O., Langbein L., Narod S.,  
RA Janin N., Goguel A.F., Bernheim A., Franke W.W., Lenoir G.M.,  
RA Feunteun J.;  
RT "Epidermolytic palmoplantar keratoderma cosegregates with a keratin 9  
mutation in a pedigree with breast and ovarian cancer.";  
Nat. Genet. 6:106-110(1994).  
[6]  
VARIANTS EPK LYS-160; GLN-162 AND TRP-162.  
MEDLINE=94214498; PubMed=7512862;  
Reis A., Hennies H.-C., Langbein L., Digweed M., Mischke D.,  
RA Dreschler M., Schroek E., Royer-Pokora B., Franke W.W., Sperling K.,  
RA Kuester W.;  
RT "Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma  
(EPK).";  
Nat. Genet. 6:174-179(1994).  
[7]  
VARIANTS EPK TRP-162 AND SER-167.  
MEDLINE=95164983; PubMed=7532199;  
Rothnagel J.A., Wojcik S., Liefer K.M., Dominey A.M., Huber M.,  
RA Hohl D., Roop D.R.;  
RT "Mutations in the IA domain of keratin 9 in patients with  
epidermolytic palmoplantar keratoderma.";  
J. Invest. Dermatol. 104:430-433(1995).  
[8]  
VARIANT EPK VAL-159.  
MEDLINE=97348961; PubMed=9204965;  
Endo H., Hatachichi A., Shinkai H.;  
RT "A novel mutation of a leucine residue in coil 1A of keratin 9 in  
epidermolytic palmoplantar keratoderma.";  
J. Invest. Dermatol. 109:113-115(1997).  
[9]  
VARIANTS EPK THR-156; VAL-156 AND GLN-162.  
MEDLINE=99072662; PubMed=9856842;  
Covello S.P., Irvine A.D., McKenna K.E., Munro C.S., Nevin N.C.,  
RA Smith F.J.D., Utito J., McLean W.H.I.;  
RT "Mutations in keratin 9 in kindreds with epidermolytic palmoplantar  
keratoderma and epidemiology in Northern Ireland.";  
J. Invest. Dermatol. 111:1207-1209(1998).



FT	CONFLICT	24	S -> F (IN REF. 2).
FT	CONFLICT	28	S -> F (IN REF. 2).
FT	CONFLICT	38	E -> L (IN REF. 2).
FT	CONFLICT	41	E -> G (IN REF. 2).
FT	CONFLICT	104	AG -> GS (IN REF. 2).
FT	CONFLICT	110	MISSING (IN REF. 2).
FT	CONFLICT	121	SY -> GC (IN REF. 2).
FT	CONFLICT	137	SY -> G (IN REF. 2).
FT	CONFLICT	148	Q -> R (IN REF. 2).
FT	CONFLICT	178	WYKKGHNSO -> VVREAROLKP (IN REF. 2).
FT	CONFLICT	263	KSDLEM -> QSVLEL (IN REF. 2).
FT	CONFLICT	284	H -> L (IN REF. 2).
FT	CONFLICT	353	E -> A (IN REF. 2).
FT	CONFLICT	394	GGRYCV -> VESLLR (IN REF. 2).
FT	CONFLICT	508	GGSHGGS -> CGSGRG (IN REF. 2).
FT	CONFLICT	523	S -> G (IN REF. 2).
FT	CONFLICT	531	H -> R (IN REF. 2).
FT	CONFLICT	534	S -> G (IN REF. 2).
FT	CONFLICT	543	S -> G (IN REF. 2).
FT	CONFLICT	547	GQ -> RR (IN REF. 2).
FT	CONFLICT	555	KS -> SGT (IN REF. 2).
ST	SEQUENCE	569 AA;	57711 MW; BEC59D4BDFE484D CRC64;
ST	SEQUENCE	55.88;	Score 144; DB 1; Length 569;
ST	Best Local Similarity	60.48;	Pred. No. 0.00018;
ST	Matches 29; Conservative	0; Mismatches 17; Indels 1;	Gaps 1;
QY	2	GGSGSGGSGGSGGSGGSGQS-	GGSGSGGSGGSGGSGGSG 47
DB	503	GGSGSGGSGGSGGSGGSGGAGHGSSGGGSGGSGGSGGSGG 550	
RESULT 7			
K1CJ_HUMAN	STANDARD;		
ID	K1CJ_HUMAN	PRT;	593 AA.
AC	P13645;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	keratin, type I cytoskeletal 10 (Cyto keratin 10) (K10) (CK 10).		
GN	KT10.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89125611; PubMed=2464696;		
RA	Rieger M., Franke W.W.;		
RT	"Identification of an orthologous mammalian cytokeratin gene. High degree of intron sequence conservation during evolution of human cytokeratin 10.";		
RL	J. Mol. Biol. 204:841-856(1988).		
RN	[2]		
RP	SEQUENCE OF 130-593 FROM N.A.		
RX	MEDLINE=88122104; PubMed=2448602;		
RA	Darmon M.Y., Senat A., Darmon M.C., Vasquez M.;		
RT	"Sequence of a cDNA encoding human keratin No 10 selected according to structural homologies of keratins and their tissue-specific expression.";		
RL	Mol. Biol. Rep. 12:277-283(1987).		
RN	[3]		
RP	SEQUENCE OF 197-593 FROM N.A.		
RX	MEDLINE=92339897; PubMed=1378806;		
RA	Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P., Kisselev L.L.;		
RT	"Exons I and VII of the gene (Ker10) encoding human keratin 10 undergo structural rearrangements within repeats.";		
RL	Gene 116:245-251(1992).		
RN	[4]		
RP	SEQUENCE OF 180-184 AND 577-589.		
RC	TISSUE=Keratinocytes;		
RX	MEDLINE=93162043; PubMed=12866667;		











between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL:	X54449;	CAA38315.1;	-;
DR	EMBL:	U40708;	AAR85863.1;	-;
DR	PIR:	S18567;	KNRZG2.	-;
KW			Cell wall; Structural protein;	Repeat; Signal.
FT	SIGNAL	1	POTENTIAL.	
FT	CHAIN	28	183	
FT	DOMAIN	28	181	
Q	SEQUENCE	183 AA;	14920 MW;	E9889452101B3893 CRC64;

Query Match 53.9%; Score 139; DB 1; Length 183;  
Best Local Similarity 48.5%; Pred. No. 0.00018;  
Matches 32; Conservative 1; Mismatches 15; Indels 18; Gaps 2;

Search completed: March 13, 2003, 15:04:05  
Job time : 2.6129 secs

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:00:59 ; Search time 4.93359 Seconds  
(without alignments)  
2088.208 Million cell updates/sec

Title: US-09-863-901-22

Perfect score: 258

Sequence: 1 SCGSGGSGGSGGSGGSGG.....SGSGGSGGSGGSGGSLRS 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	170	65.9	302	10 Q9SL09	Q9SL09 arabidopsis
2	170	65.9	738	5 Q02402	Q02402 pinctada fu
3	169	65.5	125	5 Q964C1	Q964C1 encephalito
4	168	65.1	592	16 Q9PF60	Q9PF60 xyella fas
5	166	64.3	102	5 Q964C0	Q964C0 encephalito
6	163	63.2	1422	10 Q9ZUR3	Q9ZUR3 arabidopsis
7	163	63.2	1745	10 Q9MAK1	Q9MAK1 arabidopsis
8	161.5	62.6	100	5 Q964C3	Q964C3 encephalito
9	161.5	62.6	117	5 Q964C2	Q964C2 encephalito
10	161.5	62.6	121	5 Q964C5	Q964C5 encephalito
11	161.5	62.6	138	5 Q964C4	Q964C4 encephalito
12	159.5	61.8	207	10 Q43522	Q43522 lycopersico
13	154	59.7	786	6 Q18740	Q18740 canis famil
14	153	59.3	1218	10 Q9XIH2	Q9XIH2 arabidopsis
15	150	58.1	251	3 Q9UCV7	Q9UCV7 yarrowia li
16	149	57.8	157	5 Q9GXH8	Q9GXH8 plasmodium

17	148.5	57.6	171	10	004339	004339 arabidopsis
18	147.5	57.2	87	10	Q9LRI5	Q9LRI5 pisum sativ
19	146	56.6	291	10	Q39337	Q39337 brassica na
20	146	56.6	370	5	Q9BPK8	Q9BPK8 plasmodium
21	146	56.6	370	5	Q9BPK6	Q9BPK6 plasmodium
22	145.5	56.4	118	10	Q949R3	Q949R3 arabidopsis
23	145	56.2	277	5	Q9GQP0	Q9GQP0 plasmodium
24	145	56.2	302	5	Q9GQL9	Q9GQL9 plasmodium
25	145	56.2	302	5	Q9GQL8	Q9GQL8 plasmodium
26	145	56.2	317	5	Q9BH54	Q9BH54 plasmodium
27	144.5	56.0	221	10	065514	065514 arabidopsis
28	144	55.8	561	11	Q9CXH6	Q9CXH6 mus musculu
29	143.5	55.6	208	5	Q25949	Q25949 plasmodium
30	143.5	55.6	718	12	Q91TR1	Q91TR1 tupiaia herp
31	143	55.4	588	12	Q9QEK6	Q9QEK6 cynomolgus
32	142	55.0	486	11	Q8R019	Q8R019 mus musculu
33	142	55.0	698	12	Q9QKX8	Q9QKX8 ectocarpus
34	141.5	54.8	619	12	Q91PQ9	Q91PQ9 cynomolgus
35	141	54.7	139	5	Q9GQX9	Q9GQX9 plasmodium
36	141	54.7	208	5	Q9U0A0	Q9U0A0 plasmodium
37	141	54.7	396	10	065450	065450 arabidopsis
38	140	54.3	610	5	Q9V5V8	Q9V5V8 drosophila
39	139.5	54.1	637	11	Q9D2K8	Q9D2K8 mus musculu
40	139	53.9	185	10	Q948R3	Q948R3 oryza sativ
41	138.5	53.7	271	10	Q08529	Q08529 nicotiana t
42	138.5	53.7	588	12	Q91PQ8	Q91PQ8 cynomolgus
43	138	53.5	164	5	Q9BJQ5	Q9BJQ5 plasmodium
44	138	53.5	484	16	053394	053394 mycobacteri
45	137.5	53.3	175	10	Q9LSN6	Q9LSN6 arabidopsis

#### ALIGNMENTS

#### RESULT 1

Q9SL09	ID	Q9SL09	PRELIMINARY;	PRT;	302 AA.
AC	Q9SL09;				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	At2g05580	protein.			
GN	AT2G05580				
OS	Arabidopsis thaliana	(Mouse-ear cross).			
OC	Eukaryota; Viridiplantae;	Streptophyta;			
OC	Spermatophyta; Magnoliophyta;	eudicotyledons; core eudicots; Rosidae;			
OC	eurosid3 II; Brassicales;	Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RX	MEDLINE=20083487;	PubMed=10617197;			
RA	Lin X., Kaul S., Rounsley S.D.,	Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.L.,	Barnstead M.E., Feldblyum T.V.,			
RA	Buell C.R., Ketchum K.A., Lee J.J.,	Ronning C.M., Koo H., Moffat K.S.,			
RA	Crouin L.A., Shen M., VanAken S.E.,	Umayam L., Tallon L.J., Gill J.E.,			
RA	Adams M.D., Carrera A.J., Creasy T.H.,	Goodman H.M., Somerville C.R.,			
RA	Copenhaver G.P., Preuss D., Nierman W.C.,	White O., Eisen J.A.,			
RA	Salzberg S.L., Fraser C.M., Venter J.C.;				
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana."				
RL	Nature 402:761-768(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RA	Lin X.;				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AC006220; AAC24649.1;				
DR	InterPro; IPR002952; Eggshell.				
DR	InterPro; IPR000817; PRION.				
DR	PRINTS; PR01228; EGGSHLL.				
DR	PRINTS; PR00341; PRION.				
SQ	SEQUENCE 302 AA; 26008 MW; E72A73C55825891E CRC64;				



RL	Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AC005897; AAC97230.2; -
DR	EMBL; AC007045; AAM15356.1; -
KW	Hypothetical protein.
SQ	SEQUENCE 1422 AA; 159128 MW; 0F50A356C7A608BD CRC64;
Query Match	63.2%
Best Local Similarity	61.7%
Matches	29; Conservative
Mismatches	2; Mismatches
Indels	16; Indels
Score	163; DB 10; Length 1422;
Pred. No.	4.5e-07;
Caps	0;

OS *Arabidopsis thaliana* (Mouse-ear cress).  
 OC Eukaryota: Viridiplantae: Streptophyta; Embryophyta: Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OC    EDUCATION, DIAPYCNES, DIAPYCNES, DIAPYCNES,  
OX    NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.

RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,

RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,

Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M., RA

RA Ecker J.R.:

RT	"Genomic sequence for Arabidopsis thaliana BAC F27F5 from chromosome			
RT	I.";			
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBDJ databases.			
DR	EMBL;	AC007915;	AAF69170.1;	-;
DR	InterPro;	IPR002106;	AA:RNA_LigaseII.	
DR	InterPro;	IPR000087;	Collagen.	
DR	InterPro;	IPR003653;	SUMO-protease.	
DR	Pfam;	PF02902;	Peptidase_C48;	1.
DR	PROSITE;	PS00339;	AA:RNA_LIGASE_II.2;	UNKNOWN_1.
SQ	SEQUENCE	1745 AA;	194607 MW;	26BD44B87389E2F9 CRC64;
Query Match	63.2%;	Score 163;	DB 10;	Length 1745;
Best Local Similarity	61.7%;	Score No. 5.4e-07;		

Qy 1 SGGSGGGQSGGSGGGQSGGSGGGQSGGSGGGQSG 47  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
 Db 426 SGGDEGGPSGGDEGGPSGGDEGGPSGGDEGGPN 472

Q964C3  
ID Q964C3 PRELIMINARY; PRT; 100 AA.  
AC Q964C3;

01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Spore wall protein-1 (Fragment).

OC Eukaryota; Microsporidia; Unikaryonidae; Unikaryonitooon.  
OX NCBI\_TaxID=6035;  
RN [1]

RC STRAIN=STRAIN II;  
RX MEDLINE=21270266; PubMed=11376065;  
RA Xiao L., Li L., Visvesvara G.S., Moura H., Didier E.S., Lal A.A.;

RA	Xiao L., Li L., Visvesvara G.S., Moura H., Didier E.S., Lal A.A.;
RT	"Genotyping Encephalitozoon cuniculi by Multilocus Analyses of Genes
RT	with Repetitive Sequences.";
RL	J. Clin. Microbiol. 39:2248-2253(2001).
DR	EMBL; AF340007; AK63047.1; -.
FT	NON_TER 1
FT	NON_TER 121 121
SEQ	SEQUENCE 121 AA; 9855 MW; 00F9F53A3794A221 CRC64;
Query Match 62.6%; Score 161.5; DB 5; Length 121;	
Best Local Similarity 55.4%; Pred. No. 6e-08;	
Matches 41; Conservative 1; Mismatches 5; Indels 27; Gaps	
Qy	1 SGGSGSGSGSGSG-----GGSGSGSGSG-----GGSGSGSGSG-----GGSGSGSG 36
Db	44 SDGSGSGSGSGSGSGSDGS 10
Qy	37 SG---GGSGSGSGSG 47
Db	104 SGGSGSGSGSGESG 117
RESULT 11	
ID	Q964C4 PRELIMINARY; PRT; 138 AA.
AC	O964C4;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE	Spore wall protein-1 (Fragment).
DE	Encephalitozoon cuniculi.
OC	Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX	NCBI_TaxID=6035;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CDC:V446;
RX	MEDLINE=21270266; PubMed=11376065;
RA	Xiao L., Li L., Visvesvara G.S., Moura H., Didier E.S., Lal A.A.;
RT	"Genotyping Encephalitozoon cuniculi by Multilocus Analyses of Genes
RT	with Repetitive Sequences.";
RL	J. Clin. Microbiol. 39:2248-2253(2001).
DR	EMBL; AF340008; AAK63048.1; -.
DR	PRINTS; PR01574; TUBBYPROTEIN.
FT	NON_TER 1
FT	NON_TER 138 138
SEQ	SEQUENCE 138 AA; 11123 MW; B9D4F9DA7F0A7ADF CRC64;
Query Match 62.6%; Score 161.5; DB 5; Length 138;	
Best Local Similarity 55.4%; Pred. No. 6.8e-08;	
Matches 41; Conservative 1; Mismatches 5; Indels 27; Gaps	
Qy	1 SGGSGSGSGSGSG-----GGSGSGSGSG-----GGSGSGSGSG-----GGSGSGSG 36
Db	61 SDGSGSGSGSGSGSGSDGS 121
Qy	37 SG---GGSGSGSGSG 47
Db	121 SGGSGSGSGSGESG 134
RESULT 12	
ID	Q43522 PRELIMINARY; PRT; 207 AA.
AC	Q43522;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE	Tfm5 protein.
DE	Tfm5.
GN	Lycopodium esculentum (Tomato).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC	Asteridae; euasterids I; Solanales; Solanaceae; Solanum

```

OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, UC82B; TISSUE=FRUIT;
RX MEDLINE=97201476; PubMed=9049262;
RA Santino C.G., Stanford G.L., Conner T.W.;
RT "Developmental and transgenic analysis of two tomato fruit enhanced
genes.";
RL Plant Mol. Biol. 33:405-416(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, UC82B; TISSUE=FRUIT;
RA Connor T.W.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X95262; CAA04559.1; -.
DR HSSP; P24337; 1HYP.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Try/amy1_inhbr.
DR Pfam; PF00234; try_alpha_amy1; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 207 AA; 18265 MW; A50FE71F8ED4C4D9 CRC64;

Query Match 61.8%; Score 159.5; DB 10; Length 207;
Best Local Similarity 67.3%; Pred. No. 1.5e-07;
Matches 33; Conservative 1; Mismatches 12; Indels 3; Gaps 1;

QY 2 GSGSGGSGGSGSGG---QSGGSGGSGGSGGSGGSGGSGGSGGSGG 47
||||| |||||| ||||| ||||| |||||| ||||| ||||| |||||
DB 54 GSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 102

RESULT 13
ID O18740 PRELIMINARY; PRT; 786 AA.
AC O18740;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Keratin.
GN KRT9.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Lacharme P., Hitte C., Jouquand S., Priat C., Galibert F.;
RT "Identification and analysis of the dog keratin 9 (KRT9) gene.";
RL Anim. Genet. 9:173-178(1998).
DR EMBL; AF000949; AAC26971.1; -.
DR HSSP; P10968; 2CWG.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01240; TYPEKERATIN.
DR PROSITE; PS00226; IF; UNKNOWN_1.
SQ SEQUENCE 786 AA; 76355 MW; 272AB5425DD09535 CRC64;

Query Match 59.7%; Score 154; DB 6; Length 786;
Best Local Similarity 49.3%; Pred. No. 1.6e-06;
Matches 36; Conservative 3; Mismatches 8; Indels 26; Gaps 2;

QY 1 SGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 37
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 661 SGGSSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 720

QY 38 ---GSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 47
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 721 SYGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 733

RESULT 14
QYXIH2

```

```

ID Q9XIH2 PRELIMINARY; PRT; 1218 AA.
AC Q9XIH2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE At2g16180 protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007134; AAD26959.1; -.
DR InterPro; IPR002106; AATRNA_LigaseII.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ SEQUENCE 1218 AA; 136382 MW; 4736D7FCCB72245A CRC64;

Query Match 59.3%; Score 153; DB 10; Length 1218;
Best Local Similarity 65.1%; Pred. No. 3e-06;
Matches 28; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 5 GSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSG 47
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 415 GEGSGSGGEGGEGGSGGEGGSGGEGGSGGEGGEGGEGGSGGEGG 457

RESULT 15
QYUVE7 PRELIMINARY; PRT; 251 AA.
AC QYUVE7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 21.8 kDa protein (Fragment).
GN 45C.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 20460; TISSUE=MYCELIUM;
RA Sanchez M., Dominguez A.;
RT "Gene order in a 10275 bp fragment from Yarrowia lipolytica including
YlURA5 and YlSEC65 adjacent genes conserved in four yeast species.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ006754; CAB55339.1; -.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR01574; TUBBYPROTEIN.
KW Hypothetical protein.

```

FT NON\_TER 1 1  
SQ SEQUENCE 251 AA; 21773 MW; 4BC3E20ECF45C825 CRC64;  
Query Match 58.1%; Score 150; DB 3; Length 251;  
Best Local Similarity 66.7%; Pred. No. 1.2e-06;  
Matches 34; Conservative 0; Mismatches 13; Indels 4; Gaps 2;  
QY 1 SGGSGGGQGGGGGGGG--GGSGGGQGGGGGGGG--GGSGGGQGG 47  
Db 73 SGGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 123

Search completed: March 13, 2003, 15:05:07  
Job time : 6.93359 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 14:59:19 ; Search time 58.833 Seconds  
(without alignments)  
1080.355 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGEELTGVPILVELDG.....LGFWTMPDACEKLYGGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2560	100.0	477	ABB08634	GFP fusion protein
2	2560	100.0	906	ABB08635	GFP fusion protein
3	2496.5	97.6	468	ABB08633	GFP fusion protein
4	2395.5	93.6	450	ABB08632	GFP fusion protein
5	2344	91.6	441	ABB08631	GFP fusion protein
6	2287.5	89.4	432	ABB08630	GFP fusion protein
7	1353	52.9	655	ABB75672	Protein related to
8	1295	50.6	719	AAW85012	Smad2-green floure
9	1295	50.6	719	AAW70779	EGFP-SMAD2 fusion
10	1293.5	50.5	356	ABB08620	Yeast polypeptide

11	1292.5	50.5	631	19	AAW85006	Erkl-green floures
12	1292.5	50.5	631	21	AAW70778	EGFP-Erkl fusion p
13	1291.5	50.4	1386	21	AAW85573	Hs-UNC-53/3 fragm
14	1290.5	50.4	890	21	AAW22938	GFP-HSC70 fusion p
15	1288.5	50.3	294	21	AAW22860	GFP-DEVD-annexin I
16	1288.5	50.3	294	21	AAW79638	Caspase-3 biosenso
17	1286	50.2	1070	18	AAW17789	Green fluorescent
18	1283.5	50.1	633	19	AAW85009	Jnk1-green floures
19	1283	50.1	459	21	AAW22936	GFP-HSP27 fusion p
20	1278.5	49.9	635	21	AAW85034	Green fluorescent
21	1278.5	49.9	635	21	AAW70781	EGFP-VASP fusion p
22	1277	49.9	501	18	AAW31879	GFP variants S65T
23	1276.5	49.9	1171	19	AAW85037	Green fluorescent
24	1274	49.8	250	22	AAW66492	Green fluorescence
25	1272	49.7	558	19	AAW48662	GR fusion protein
26	1271.5	49.7	1090	21	AAW85576	Hs-UNC-53/1 fragm
27	1271	49.6	397	20	AAW42176	EGFP/DRM fusion pr
28	1271	49.6	429	20	AAW42175	EGFP/DRM fusion pr
29	1271	49.6	501	18	AAW31877	GFP variants S65C
30	1271	49.6	501	18	AAW31878	GFP variants S65C
31	1271	49.6	514	18	AAW31876	GFP variants S65C
32	1270	49.6	238	18	AAW22101	Aequorea victoria
33	1270	49.6	238	20	AAW96328	Humanised green fl
34	1270	49.6	247	20	AAW96329	Humanised green fl
35	1269.5	49.6	359	20	AAW42179	EGFP/DRM fusion pr
36	1269.5	49.6	359	20	AAW42180	EGFP/DRM fusion pr
37	1269.5	49.6	391	20	AAW42178	EGFP/DRM fusion pr
38	1269.5	49.6	997	19	AAW85032	Green fluorescent
39	1269	49.6	607	19	AAW85010	p38-green flouresc
40	1269	49.6	783	21	AAW22939	GFP-HSF1 fusion pr
41	1268	49.5	248	23	AAW68319	Jellyfish green fl
42	1268	49.5	403	20	AAW42177	EGFP/DRM fusion pr
43	1268	49.5	842	19	AAW85008	Grk5-green floures
44	1267	49.5	605	19	AAW85007	Erk2-green floures
45	1267	49.5	727	19	AAW85041	Green fluorescent

ALIGNMENTS

.... RESULT 1

ABB08634

ID ABB08634 standard; Protein: 477 AA.

XX

AC ABB08634;

XX

DT 03-MAY-2002 (first entry)

XX

DE GFP fusion protein G5A SEQ ID NO 5.

XX

KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;

KW Chemiluminescence Resonance Energy Transfer; CREt;

KW central nervous system; neural network.

XX

OS Synthetic.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Key Location/Qualifiers

Misc-difference 445

/note= "Encoded by AQT"

WO200192300-A2.

PD 06-DEC-2001.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX







fluorescent molecule covalently linked with a photoprotein, useful for monitoring calcium fluxes or for detecting electrical activity in a group of neural cells -

Claim 2; Page 30; 61pp; English.

The invention relates to a fusion protein (ABB08630-ABB08635) for energy transfer from aequorin to green fluorescent protein by chemiluminescence resonance energy transfer (CRET), which comprises a fluorescent molecule covalently linked with a photoprotein. The fusion protein comprises the formula: GFP - LINKER - AEQ

GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a polypeptide of 4-63 amino acids.

The fusion protein is useful as a bioluminescent Ca<sup>2+</sup> reporter at the single cell level. The fusion protein, composition or bioluminescent system is useful for monitoring calcium fluxes in real time. This is particularly useful for understanding the development, the plasticity and the functioning of the central nervous system. The fusion protein comprising the bioluminescent system is useful for detecting electrical activity in a group of neural cells, for making it possible to complete the phenotype study of mutants, or for observing the calcium activity in a population of connected cells, for example in a neural network.

XX

SQ Sequence 432 AA;

Query Match 89.4%; Score 2287.5; DB 23; Length 432;  
Best Local Similarity 90.4%; Pred. No. 1.2e-184;  
Matches 431; Conservative 0; Mismatches 1; Indels 45; Gaps 1;

QY 1 MSKGEELFTGVPTLVELDGVNGHKFSVSGEGDATYKGLTLKFKTCTTGKLPVPMPTL 60  
DB 1 MSKGEELFTGVPTLVELDGVNGHKFSVSGEGDATYKGLTLKFKTCTTGKLPVPMPTL 60

QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSPMPGYVQERTIFFKDDGNYKTRAEVFEFGDTLV 120  
DB 61 VTTLTYGVQCFSRYPDHMKQHDFFKSPMPGYVQERTIFFKDDGNYKTRAEVFEFGDTLV 120

QY 121 NRIELKGIDFKEGNILGHKLEYNYNHNVYINADKQKNGIKANFKIRHNIEDGSVQLAD 180  
DB 121 NRIELKGIDFKEGNILGHKLEYNYNHNVYINADKQKNGIKANFKIRHNIEDGSVQLAD 180

QY 181 HYQONTPIGDPVLLPNHLYLTQSALSKDPNEKRDMHVLLEFVTAAGITGHGDELYKSG 240  
DB 181 HYQONTPIGDPVLLPNHLYLTQSALSKDPNEKRDMHVLLEFVTAAGITGHGDELYK-- 238

QY 241 GSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRSLTSDFDNPRW 300  
DB 239 -----SGLRSVKLTSDFDNPRW 255

QY 301 IGRHKHMFNLVDVNHNGKISLDEMVKASDIVNNLGATPEQAKRHRKDAVEAFEGGAGMK 360  
DB 256 IGRHKHMFNLVDVNHNGKISLDEMVKASDIVNNLGATPEQAKRHRKDAVEAFEGGAGMK 315

QY 361 YGVETDMPAYIEGKKIATDELEKYAKNEPTLIRWGDAFLDIVDKDQNGAITLDEWKAY 420  
DB 316 YGVETDMPAYIEGKKIATDELEKYAKNEPTLIRWGDAFLDIVDKDQNGAITLDEWKAY 375

QY 421 TKAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGEFWYTMDPACEKLYGAVP 477  
DB 376 TKAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGEFWYTMDPACEKLYGAVP 432

RESULT 7  
ABB76672

ID ABB76672 standard; protein; 655 AA.

XX ABB76672;

XX

DT 03-SEP-2002 (first entry)

XX Protein related to Bombyx mori silk fibroin.

DE

XX Silk; fibroin; textile industry.

XX OS Unidentified.  
 XX XX WO200240528-A1.  
 XX PN 23-MAY-2002.  
 XX PD 26-OCT-2001; 2001WO-CN01506.  
 XX PF 26-OCT-2000; 2000CN-0125859.  
 XX PR (SHAN-) SHANGHAI BIO-CHEM INST CHINESE ACAD SCI.  
 XX PA Lu C, Huang J, Zhao Y, Zhang F, Chen X;  
 XX PI WPI; 2002-427084/45.  
 XX DR Method for producing non-natural silk by Bombyx mori with modifying  
 XX PT heavy and light chains of fibroin in natural silk by DNA recombination  
 XX PT technology and protein engineering, applicable in sericulture and  
 XX PT textile industry -  
 XX PS Claim 4; Page 40-42; 50pp; Chinese.  
 XX CC This invention relates to a silk which constitutes recombinant Bombyx  
 XX CC mori silk fibroin obtained by modification or variation at the heavy  
 XX CC or light chains of the fibroin. The method is for producing non-natural  
 XX CC silk, which is applicable in sericulture and textile industry. Such  
 XX CC modified silk has improved performance. The present sequence is a  
 XX CC protein related to the invention.  
 XX XX  
 XX XX Sequence 655 AA;  
 XX  
 Query Match 52.9%; Score 1353; DB 23; Length 655;  
 Best Local Similarity 86.0%; Pred. No. 1.7e-105;  
 Matches 264; Conservative 4; Mismatches 19; Indels 20; Gaps 5;  
 QY 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVWPPTL 60  
 DB 6 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVWPPTL 65  
 QY 61 VTTLTYGVOCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 120  
 DB 66 VTTLTYGVOCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 125  
 QY 121 NRLEKIDFKEDGNILGHKLEYNHSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 180  
 DB 126 NRLEKIDFKEDGNILGHKLEYNHSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 185  
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTGHMDELYK-- 238  
 DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTGHMDELYKGI 245  
 QY 239 --SGSGSG--GGSGSGSG-----SGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 281  
 DB 246 PADGRSGGLGGGPGGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304  
 QY 282 GSGGLRS 288  
 DB 305 GTGGLGS 311  
 RESULT 8  
 ID AAW85012 standard; Protein; 719 AA.  
 AC AAW85012;  
 XX 08-FEB-1999 (first entry)  
 XX Smad2-green fluorescent protein fusion product.  
 DE Human; Smad2 gene; fusion protein; green fluorescent protein; GFP;  
 KW

KW intracellular signalling; chimera.  
 XX Chimeric - Aequorea victoria.  
 OS Chimeric - Homo sapiens.  
 XX WO9845704-A2.  
 XX PD 15-OCT-1998.  
 XX PF 07-APR-1998; 98WO-DK00145.  
 XX PR 07-APR-1997; 97DK-0000392.  
 XX PA (NOVO ) NOVO-NORDISK AS.  
 XX PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O;  
 PI Tullin S;  
 XX WPI; 1998-594491/50.  
 DR N-PSDB; AAV71027.  
 XX  
 PT Determining effect on signalling pathways in live cells from  
 PT redistribution of luminophores - specifically fusions of green  
 PT fluorescent protein with a signalling component, and new apparatus,  
 PT particularly for identifying toxins and potential therapeutic agents  
 PS Example 5; Pages 102-103; 326pp; English.  
 XX  
 CC The present sequence represents a human Smad2-green fluorescent  
 CC protein fusion product. The fusion protein is used in an assay  
 CC that exemplifies the invention. The specification describes how  
 CC quantitative information about the influence of a molecule on a cellular  
 CC response is obtained by recording the variation, caused by the molecule,  
 CC on mechanically intact living cells, in the spatially distributed light  
 CC emitted from a luminophore present in the cells. The variation in light  
 CC distribution is processed to provide information that correlates spatial  
 CC distribution to the degree of the molecule. The method is used to  
 CC identify agents that (in)directly affect intracellular signalling,  
 CC especially to screen for potential therapeutic agents or toxins, and  
 CC to identify new drug targets.  
 XX  
 XX Sequence 719 AA;  
 Query Match 50.6%; Score 1295; DB 19; Length 719;  
 Best Local Similarity 81.0%; Pred. No. 1.5e-100;  
 Matches 251; Conservative 9; Mismatches 36; Indels 14; Gaps 1;  
 QY 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVWPPTL 60  
 DB 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVWPPTL 61  
 QY 61 VTTLTYGVOCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 120  
 DB 62 VTTLTYGVOCFSRYPDHMKQHDFFKSAPEGVYQERTIFFKDDGNKTRAEVKFEGDTLV 121  
 QY 121 NRLEKIDFKEDGNILGHKLEYNHSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 180  
 DB 122 NRLEKIDFKEDGNILGHKLEYNHSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 181  
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTGHMDELYKSG 240  
 DB 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTGHMDELYKSG 241  
 QY 241 GSG 286  
 DB 242 LRSRAQASNTMSSILPFTPPVVKLLGWKKKSGAGGAGGGEQKQBEKWKCAVKSUV 301  
 QY 287 RSVKLTSDFD 296  
 DB 302 KKLKKTGRLD 311  
 RESULT 9

AAV70779

ID AAV70779 standard; Protein: 719 AA.  
XX  
AC AAV70779;  
XX  
DT 31-JUL-2000 (first entry)  
XX  
DE EGFP-SMAD2 fusion protein construct.  
XX  
KW Fusion construct; EGFP-SMAD2; Green fluorescent protein; cytokine;  
KW cellular response; luminophore; screening; medicament; signal transducer;  
KW light/fluorescence intensity; intracellular signalling pathway; mutant;  
KW fluorescent probe; signal transduction; transforming growth factor-beta;  
KW TGF-beta.  
XX  
OS Chimeric - Homo sapiens.  
OS Chimeric - Aequorea victoria.  
XX  
PN WO200023615-A2.  
XX  
PD 27-APR-2000.  
XX  
PF 15-OCT-1999; 99WO-DK00562.  
XX  
PR 15-OCT-1998; 98DK-0001320.  
XX  
PA (BIOI-) BIOIMAGE AS.  
XX  
PI Arkhammar POG, Terry BR, Scudder KM, Bjorn SP, Thastrup O;  
PI Hagel G;  
XX  
DR WPI: 2000-339710/29.  
DR N-PSDB: AAD00037.  
XX  
XX Quantifying cellular response to an influence, useful for detecting  
PT intracellular translocation or redistribution of biologically active  
PT substances comprises recording changes in spatially distributed light  
PT emitted from a luminophore -  
XX  
PS Example 4; Page 110-112; 150pp; English.  
XX  
CC The patent discloses a method for extracting quantitative information  
CC relating to an influence on a cellular response in mechanically intact  
CC living cells, by recording variation in spatially distributed light  
CC emitted from a luminophore, as a change in light/fluorescence intensity.  
CC The luminophore present in the cells, is capable of being redistributed  
CC in a manner which is related with the degree of the influence. This  
CC method is useful as a screening program, for the identification of a  
CC biologically active substance, that directly or indirectly affects an  
CC intracellular signalling pathway. It is also potentially useful as a  
CC medicament. The fluorescent probe is useful in the back-tracking of  
CC signal transduction pathways. The present sequence is a fusion protein  
CC construct EGFP-SMAD2, comprising the human Smad2 gene, fused to a  
CC derivative of the luminiscent green fluorescent protein, EGFP. Smad2 is  
CC a signal transducer, that is induced by some members of the transforming  
CC growth factor-beta (TGF) family of cytokines. This construct is under the  
CC control of a CMV promoter and functions as the luminophore, that is  
CC useful to monitor signalling pathways and identify compounds that  
CC modulate the pathways in living cells.  
XX  
SQ Sequence 719 AA;

Query Match 50.6%; Score 1295; DB 21; Length 719;  
Best Local Similarity 81.0%; Pred. No. 1.5e-100;  
Matches 251; Conservative 9; Mismatches 36; Indels 14; Gaps 1;  
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLKTLKFICTTGKLPVPWPTL 60  
DB 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLKTLKFICTTGKLPVPWPTL 61  
QY 61 VTTLTGVQCFSRYPDMKQKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFECDTLV 120  
DB 62 VTTLTGVQCFSRYPDMKQKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFECDTLV 121

QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHINIEDSGVOLAD 180  
DB NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHINIEDSGVOLAD 181  
QY 181 HYQONTPTGDPVLLPDNHYLSTQSALSADPNKRDHNVLLFEVTAAGTTHGMDELYKSG 240  
DB HYQONTPTGDPVLLPDNHYLSTQSALSADPNKRDHNVLLFEVTAAGTTHGMDELYKSG 241  
QY 241 GSGSGGSG 286  
DB LRSRAQASNSTWSSILPFTPPVVKLLGHWKKSAGSGGAGGGEQNGQEKWCEKAVKSLV 301  
QY 287 RSVKLTSDFD 296  
DB 302 KKLKKTGRLD 311  
RESULT 10  
ABB08620  
ID ABB08620 standard; Protein: 356 AA.  
XX  
AC ABB08620;  
XX  
DT 10-APR-2002 (first entry)  
XX  
DE Yeast polypeptide 1.  
XX  
KW Yeast; drug; cell wall; GPI anchor protein.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN WO200183733-A1.  
XX  
PD 08-NOV-2001.  
XX  
PF 26-APR-2001; 2001WO-JP03630.  
XX  
PR 01-MAY-2000; 2000JP-0132041.  
XX  
PA (DAUC ) DAIICHI PHARM CO LTD.  
XX  
PI Kitamura A, Someya K, Nakajima R;  
XX  
DR WPI: 2002-097496/13.  
DR N-PSDB: ABA97833.  
XX  
PT Screening for drugs that act on cell walls, involves culturing  
PT microorganisms with a reporter protein acting as a GPI anchor protein  
PT in their cell walls -  
XX  
PS Disclosure; Fig 4-5; 44pp; Japanese.  
XX  
CC The invention relates to screening for drugs that act on cell walls,  
CC comprising culturing microorganisms with a reporter protein acting as a  
CC GPI anchor protein in their cell walls in the presence of the test  
CC substance and assaying the amount of reporter protein produced in the  
CC culture. The present sequence is that of a polypeptide useful to the  
CC invention.  
XX  
SQ Sequence 356 AA;

Query Match 50.5%; Score 1293.5; DB 23; Length 356;  
Best Local Similarity 88.0%; Pred. No. 7.8e-101;  
Matches 250; Conservative 3; Mismatches 26; Indels 5; Gaps 2;  
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLKTLKFICTTGKLPVPWPTL 60  
DB 40 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYGLKTLKFICTTGKLPVPWPTL 99  
QY 61 VTTLTGVQCFSRYPDMKQKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFECDTLV 120  
DB 100 VTTLTGVQCFSRYPDMKQKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFECDTLV 159









PN WO200050872-A2.  
XX 31-AUG-2000.  
XX 25-FEB-2000; 200WO-US04794.  
XX 26-FEB-1999; 99US-0122152.  
PR 08-MAR-1999; 99US-0123399.  
PR 12-JUL-1999; 99US-0352171.  
XX (CELL-) CELLOMICS INC.  
XX Giuliano KA, Kapur R;  
XX WPI; 2000-594086/56.  
DR N-PSDB; AAA93351.  
XX Automated cell-based characterization of toxin by contacting cells  
PT containing luminescent reporter molecules with test substance and  
PT analyzing optically -  
XX  
PS Example 11; Page 173-174; 336pp; English.  
XX  
CC The invention relates to systems, methods and reagents for cell-based  
CC screening or detection of compounds which affect particular biological  
CC functions. The methods of the invention utilise fluorescent biodeceptor  
CC molecules which, when acted on by a compound of interest, cause an  
CC alteration in the cellular distribution of at least the fluorescent  
CC moiety. In one embodiment, the biosensors comprise heat shock proteins  
CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent  
CC protein (GFP), or derivatives thereof). Such biosensors are located in  
CC the cytoplasm, but on stress activation translocate to the nucleus. In  
CC another embodiment biodeceptor proteins can be used to detect protease  
CC activity. Such protease biodeceptor fusion proteins comprise one or more  
CC fluorescent proteins; a recognition signal which is cleaved by the  
CC protease; and at least one cellular localisation signal. The latter two  
CC components may be components of a single protein which is acted upon by  
CC the protease, or may be from heterologous sources. Due to the  
CC localisation signal, the biodeceptor protein is localised to a  
CC particular region of the cell. Once acted on by the protease of interest,  
CC the fluorescent protein is cleaved from the localisation sequence, and  
CC is free to migrate to other locations within the cell. The presence of a  
CC second localisation signal attached to the fluorescent protein enables  
CC the fluorescent protein to be directed to a different cellular  
CC compartment after cleavage of the protease recognition sequence. The  
CC change in distribution of the fluorescent protein can be detected using  
CC imaging methods with a high degree of spatial resolution. The methods  
CC and biosensors of the invention can be used to investigate a wide range  
CC of cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased  
CC quantity and speed of data collection and faster evaluation of drug  
CC candidates. Sequences AAB22860-B22876 and AAB22936-B22941 represent  
CC biosensor fusion proteins produced in an exemplification of the  
CC invention.  
XX  
SQ Sequence 294 AA;  
Query Match 50.3%; Score 1288.5; DB 21; Length 294;  
Best Local Similarity 91.0%; Pred. No. 1.6e-100;  
Matches 244; Conservative 6; Mismatches 17; Indels 1; Gaps 1;  
QY 1 MSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGLTKLFCICTGKLPVPWPTL 60  
DB :|||||  
2 VSKGEELFTGVVPIVLVELDGVNGHKFSVSGEGDATYGLTKLFCICTGKLPVPWPTL 61  
QY 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFGDTLV 120

Db 62 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFGDTLV 121  
QY 121 NRIELKGDIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKANFKIRHINIEDGSGVOLAD 180  
Db 122 NRIELKGDIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKANFKIRHINIEDGSGVOLAD 181  
QY 181 HYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITTHGMDELYKSG 240  
Db 182 HYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITTHGMDELYKSG 241  
QY 241 GSISGGSGGSGSGG-QSGSGSGSGGSG 267  
Db 242 LRSGAGAGAGAGAGADEVDGAGADEVDG 269  
Search completed: March 13, 2003, 15:03:39  
Job time : 61.833 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:02:34 ; Search time 22.6281 Seconds  
(without alignments)  
620.235 Million cell update

**Title:** US-09-863-901-5

Perfect score:

Sequence: 1 MSKGEELFTGVPIVLVDG.....LGFWYTMDPACEKLYGGAVP 477

Scoring table: BLOSUM62

Learning curve: 0.000002  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Minimum	DB seq	length:
Maximum	DB seq	length: 2000000000

\*\*\*\*\*  
 DD SEC TENCH: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

## Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cqn2\_6/ptodata/2/laa/5A\_COMB.pcp.\*

2: /cgn2\_6/ptodata/2/laa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pcp:\*

5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pe

6: /cgn2\_6/ptodata/2/laa/backfiles1.pe

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	1290.5	50.4	890	4	US-09-513-783A-174	Sequence 174, Appl
2	1288.5	50.3	294	4	US-09-513-783A-2	Sequence 2, Appl
3	1286	50.2	1070	4	US-09-091-042A-2	Sequence 2, Appl
4	1283	50.1	459	4	US-09-513-783A-170	Sequence 170, Appl
5	1270	49.6	238	3	US-08-893-327-16	Sequence 16, Appl
6	1270	49.6	247	3	US-08-893-327-18	Sequence 18, Appl
7	1269	49.6	783	4	US-09-513-783A-176	Sequence 176, Appl
8	1267	49.5	941	4	US-09-513-783A-172	Sequence 172, Appl
9	1286	49.5	1452	4	US-09-127-227-2	Sequence 2, Appl
10	1265.5	49.4	350	4	US-09-513-783A-30	Sequence 30, Appl
11	1265	49.4	805	4	US-09-513-783A-178	Sequence 178, Appl
12	1265	49.4	1407	4	US-08-974-549A-628	Sequence 628, Appl
13	1263	49.3	302	4	US-09-513-783A-18	Sequence 18, Appl
14	1262	49.3	238	1	US-08-337-915A-2	Sequence 2, Appl
15	1262	49.3	238	4	US-09-121-539-1	Sequence 1, Appl
16	1262	49.3	238	5	PCR-US95-14692-2	Sequence 2, Appl
17	1259.5	49.2	642	2	US-08-818-253-2	Sequence 2, Appl
18	1259.5	49.2	642	4	US-08-818-253-2	Sequence 2, Appl
19	1259.5	49.2	652	2	US-08-818-253-4	Sequence 4, Appl
20	1259.5	49.2	652	4	US-08-818-252-4	Sequence 4, Appl
21	1259	49.2	295	4	US-09-513-783A-20	Sequence 20, Appl
22	1258	49.1	238	1	US-08-753-143-2	Sequence 2, Appl
23	1258	49.1	238	2	US-08-679-865-2	Sequence 2, Appl
24	1258	49.1	238	2	US-08-680-876-2	Sequence 2, Appl
25	1258	49.1	238	2	US-08-792-553-2	Sequence 2, Appl
26	1258	49.1	238	3	US-08-753-144-2	Sequence 2, Appl
27	1258	49.1	238	4	US-09-094-359-2	Sequence 2, Appl



```
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170

Query Match          50.1%; Score 1283; DB 4; Length 459;
Best Local Similarity 66.8%; Pred No. 8.5e-100;
Matches 266; Conservative 11; Mismatches 55; Indels 56; Gaps 6;

QY 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPTL 60
Db 2 VSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPTL 61
QY 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 62 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKANFKIRHNIEDGSVQLAD 180
Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKANFKIRHNIEDGSVQLAD 181
QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYKSG 240
Db 182 HYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYKSG 241
QY 241 GSGSGGSGGSGS-----GGQS-----G 258
Db 242 LRSRAASRAESAMTERRVPFSLLRGPSWDPFRDWPYHSRFLDQAFGLPRLPEESQWLIG 301
QY 259 GSGSGG-----OSGSGSGSGSGSGSGSGSLRSLVSKLTSDFDNPRWIGRKHMF 308
Db 302 GSWPGYVRPLPPAIESPAVAAPAYSRALSRLQSLSGSVSEIRHTAD-----RWRVS----- 352
QY 309 NPLDVNHNGKISLDEMYKASDIVINNLGATPEQAKRH 346
Db 353 --LDVNH---FAPDELTVKTKDGVVEITGKHEERODEH 385

RESULT 5
US-08-893-327-16
; Sequence 16, Application US/08893327
; Patent No. 6020192
; GENERAL INFORMATION:
; APPLICANT: Zolotukhin, Sergei
; APPLICANT: Hauswirth, William W.
; APPLICANT: Muzyczka, Nicholas
; TITLE OF INVENTION: Humanized Green Fluorescent Protein
; TITLE OF INVENTION: Genes and Methods
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,327
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/588,201
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UFLA:062\KIT
; TELEPHONE: (512) 418-3000
; MOLECULE TYPE: protein
```

```
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-893-327-16

Query Match          49.6%; Score 1270; DB 3; Length 238;
Best Local Similarity 99.6%; Pred No. 4.2e-99;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPTL 60
Db 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGEDATYGKLTLPKFICTTGKLPVPWPTL 60
QY 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKANFKIRHNIEDGSVQLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIIMADKQNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYK 238
Db 181 HYQONTPTIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYK 238

RESULT 6
US-08-893-327-18
; Sequence 18, Application US/08893327
; Patent No. 6020192
; GENERAL INFORMATION:
; APPLICANT: Zolotukhin, Sergei
; APPLICANT: Hauswirth, William W.
; APPLICANT: Muzyczka, Nicholas
; TITLE OF INVENTION: Humanized Green Fluorescent Protein
; TITLE OF INVENTION: Genes and Methods
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,327
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/588,201
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UFLA:062\KIT
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```





```
Db 1275 VTTFTYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 1334
Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 1335 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 1394
Qy 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITHGMDELYK 238
Db 1395 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITHGMDELYK 1452

RESULT 10
US-09-513-783A-30
; Sequence 30, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 30
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; NLS-Fred25-cellubrevin construct
US-09-513-783A-30

Query Match 49.4%; Score 1265.5; DB 4; Length 350;
Best Local Similarity 93.4%; Pred. No. 1.7e-98;
Matches 239; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Qy 2 SKGEELFTGVVPILVELDGVNKGHFVS GEGEGDATYKLT LKFTCTTGKLPVPMPTLV 61
Db 9 SKGEELFTGVVPILVELDGVNKGHFVS GEGEGDATYKLT LKFTCTTGKLPVPMPTLV 68
Qy 62 TLTLYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db 69 TLTLYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 128
Qy 122 RIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVOLADH 181
Db 129 RIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVOLADH 188
Qy 182 YQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITHGMDELYKSGG 241
Db 189 YQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITHGMDELYNTCM 248
Qy 242 SGSGGSGSGSGSGGS 257
Db 249 S-TGVPSSGSAATGSN 263

RESULT 11
US-09-513-783A-178
; Sequence 178, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 178
; LENGTH: 805
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-NFKB
US-09-513-783A-178

Query Match 49.4%; Score 1265; DB 4; Length 805;
Best Local Similarity 98.8%; Pred. No. 6.1e-98;
Matches 237; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVVPILVELDGVNKGHFVS GEGEGDATYKLT LKFTCTTGKLPVPMPTLV 60
Db 2 VSKGEELFTGVVPILVELDGVNKGHFVS GEGEGDATYKLT LKFTCTTGKLPVPMPTLV 61
Qy 61 VTTLYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 62 VTTLYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 181
Qy 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITHGMDELYKSG 240
Db 182 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGITHGMDELYKSG 241

RESULT 12
US-08-974-549A-628
; Sequence 628, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
```

0;

RESULT 13

ATTORNEY/AGENT INFORMATION: 435  
NAME: Spitals, John P.  
REGISTRATION NUMBER: 29,215  
REFERENCE/DOCKET NUMBER: 1279-178  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 977-1001  
TELEFAX: (213) 977-1003  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-337-915A-2

Query Match 49.38; Score 1263; DB 4; Length 302;  
Best Local 88.66; Score 1263; DB 4; Length 302;  
Best Global 88.66; Score 1263; DB 4; Length 302;

RESULT 14

US-08-337-915A-2  
: Sequence 2, Application US/08337915A

GENERAL INFORMATION:  
 APPLICANT: Tsieng, Roger Y.  
 APPLICANT: Helm, Roger  
 TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Robbins, Berliner & Carson  
 STREET: 201 No. 5625048th Figueroa Street, Suite 500  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: USA  
 ZIP: 90012  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/337,915A  
 FILING DATE:

ATTORNEY/AGENT INFORMATION: 435  
NAME: Spitals, John P.  
REGISTRATION NUMBER: 29,  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 977-1003  
TELEFAX: (213) 977-1003  
INFORMATION FOR SEQ ID NO: 2  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLLECULE TYPE: protein  
US-08-337-915A-2

Query Match 49.3%; Score 1262; DB 1; Length 238;  
Best Local Similarity 98.7%; Pred. No. 2e-98;  
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFLICTTGKLPVPWPTL 60  
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFLICTTGKLPVPWPTL 60

QY 61 VTTLTYGVQCFSPYDPMKQHDFFKSPMPGEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120  
DB 61 VTTFSGVQCFSPYDPMKQHDFFKSPMPGEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120

QY 121 NRLEKGIQDFKEDGNILGHKLEYNSHNVIYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180  
DB 121 NRLEKGIQDFKEDGNILGHKLEYNSHNVIYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180

QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 15  
US-09-121-539-1  
: Sequence 1, Application US/09121539B  
: Patent No. 6194548  
: GENERAL INFORMATION:  
: APPLICANT: Osumi, Takashi  
: APPLICANT: Tsukamoto, Toshio  
: APPLICANT: Tsukamoto, No. 61945481yo  
: APPLICANT: Yamasaki, Masatoshi  
: TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT  
: FILE REFERENCE: 046124-5005  
: CURRENT APPLICATION NUMBER: US/09/121,539B  
: CURRENT FILING DATE: 1998-07-24  
: PRIOR APPLICATION NUMBER: JP 026418/1998  
: PRIOR FILING DATE: 1998-01-23  
: NUMBER OF SEQ ID NOS: 14  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 1  
: LENGTH: 238  
: TYPE: PRT  
: ORGANISM: Aequorea victoria  
: FEATURE:  
: OTHER INFORMATION: Green fluorescent protein  
US-09-121-539-1

Query Match 49.3%; Score 1262; DB 4; Length 238;  
Best Local Similarity 98.7%; Pred. No. 2e-98;  
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFLICTTGKLPVPWPTL 60  
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYKGLTLKFLICTTGKLPVPWPTL 60

QY 61 VTTLTYGVQCFSPYDPMKQHDFFKSPMPGEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120  
DB 61 VTTFSGVQCFSPYDPMKQHDFFKSPMPGEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120

QY 121 NRLEKGIQDFKEDGNILGHKLEYNSHNVIYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180  
DB 121 NRLEKGIQDFKEDGNILGHKLEYNSHNVIYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180

QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDELYK 238



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:05:14 ; Search time 19.9127 Seconds  
(without alignments)  
1104.115 Million cell updates/sec

Title: US-09-863-901-5  
Perfect score: 2560  
Sequence: 1 MSKGEELFTGVVPILVELDG.....LGFWYTMDPACKLYGGAVP 477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues  
Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	49.5	1452	12	US-10-050-673-2
2	1262	49.3	238	9	US-09-866-538-2
3	1262	49.3	238	9	US-09-900-345A-125
4	1262	49.3	238	10	US-09-920-922-4
5	1262	49.3	238	10	US-09-852-000-1
6	1259.5	49.2	642	9	US-09-554-000-2
7	1259.5	49.2	652	9	US-09-554-000-4
8	1259	49.2	243	9	US-09-900-345A-60
9	1259	49.2	243	9	US-09-900-345A-62
10	1259	49.2	243	9	US-09-900-345A-64
11	1259	49.2	243	9	US-09-900-345A-66
12	1259	49.2	243	9	US-09-900-345A-68
13	1259	49.2	243	9	US-09-900-345A-70
14	1258	49.1	238	9	US-10-057-505-2
15	1258	49.1	238	10	US-09-884-681-2
16	1258	49.1	238	12	US-10-024-686-2
17	1258	49.1	243	9	US-09-900-345A-54
18	1258	49.1	243	9	US-09-900-345A-56
19	1258	49.1	243	9	US-09-900-345A-58

ALIGNMENTS

RESULT 1  
US-10-050-673-2  
; Sequence 2, Application US/10050673  
; Patent No. US20020151033A1  
; GENERAL INFORMATION:  
; APPLICANT: David M. Knipe  
; APPLICANT: Travis J. Taylor  
; APPLICANT: Elizabeth McNamee  
; TITLE OF INVENTION: Replication-Competent Virus Expressing A  
; FILE REFERENCE: H098-05  
; CURRENT APPLICATION NUMBER: US/10/050.673  
; CURRENT FILING DATE: 2002-01-16  
; PRIOR FILING DATE: 1998-07-31  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1452  
; TYPE: PFT  
; ORGANISM: herpesvirus  
US-10-050-673-2

Query Match	49.5%	Score 1266;	DB 12;	Length 1452;
Best Local Similarity	99.2%	Pred. No. 3e+80;		
Matches	236;	Conservative	0;	Mismatches 2;
Indels	0;	Gaps	0;	
Qy	1	MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYTKLTKFICTTGKLPVWPTL	60	
Db	1215	MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYTKLTKFICTTGKLPVWPTL	1274	
Qy	61	VTTLTGYQCFSRYPDHMKQHDFFKSAPEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV	120	
Db	1275	VTTFTYGVQCFSRYPDHMKQHDFFKSAPEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV	1334	
Qy	121	NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGKANKFKIRHNIEDGSVQLAD	180	
Db	1335	NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGKANKFKIRHNIEDGSVQLAD	1394	
Qy	181	HYQQNTPDGDPVLLPDNNHLYSTQSAISKDPNEKRDHVMVLLFEFVTAAGITHGMDELYK	238	
Db	1395	HYQQNTPDGDPVLLPDNNHLYSTQSAISKDPNEKRDHVMVLLFEFVTAAGITHGMDELYK	1452	

```
RESULT 2
US-09-866-538-2
; Sequence 2, Application US/09866538
; Publication No. US20030032089A1
; GENERAL INFORMATION:
; APPLICANT: TSIEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-866-538-2

Query Match      49.3%; Score 1262; DB 9; Length 238;
Best Local Similarity 98.7%; Pred. No. 6.9e-81;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPTL 60
QY 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 3
US-09-900-345A-125
; Sequence 125, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-50S
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP humanized
; US-09-900-345A-125

Query Match      49.3%; Score 1262; DB 9; Length 238;
Best Local Similarity 98.7%; Pred. No. 6.9e-81;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPTL 60
QY 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 4
US-09-920-922-4
; Sequence 4, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-4

Query Match      49.3%; Score 1262; DB 10; Length 238;
Best Local Similarity 98.7%; Pred. No. 6.9e-81;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGDATYGLTLKFKICTTGKLPVPWPTL 60
QY 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
DB 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 5
US-09-852-000-1
; Sequence 1, Application US/09852000
; Patent No. US20020099170A1
; GENERAL INFORMATION:
; APPLICANT: Osumi, Takashi
; APPLICANT: Tsukamoto, Toshio
; APPLICANT: Tsukamoto, No. US20020099170A1yo
; APPLICANT: Yamasaki, Masatoshi
; TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: 046124-5005-US
; CURRENT APPLICATION NUMBER: US/09/852,000
```

	;	CURRENT FILING DATE:	2001-05-10	
	;	PRIOR APPLICATION NUMBER:	JP 026418/1998	
	;	PRIOR FILING DATE:	1998-01-23	
	;	PRIOR APPLICATION NUMBER:	US 09/121,539	
	;	PRIOR FILING DATE:	1998-07-24	
	;	PRIOR APPLICATION NUMBER:	US 09/615,655	
	;	PRIOR FILING DATE:	2000-07-13	
	;	NUMBER OF SEQ ID NOS:	14	
	;	SOFTWARE:	PatentIn Ver. 2.0	
	;	SEQ ID NO 1		
	;	LENGTH:	238	
	;	TYPE:	PRT	
	;	ORGANISM:	Aequorea victoria	
	;	FEATURE:		
	;	OTHER INFORMATION:	Green fluorescent protein	
	;	US-09-852-000-1		
		Query Match	49.3%; Score 1262; DB 10; Length 238;	
		Best Local Similarity	98.7%; Pred. No. 6.9e-81;	
		Matches 235; Conservative	1; Mismatches 2; Indels 0; Gaps 0;	
QY	1	MSKGEELFTGVVPILVELDGNVGHKFSVS	GEGDATYGKLTFLKICITGKLPVPWPTL 60	
DB	1	MSKGEELFTGVVPILVELDGNVGHKFSVS	GEGDATYGKLTFLKICITGKLPVPWPTL 60	
QY	61	VTTLTGYQCFSRYPDHMKQHDFPKSAMPE	GYVQERTIFFKDGNKYTKRAEVKEFGDTLV 120	
DB	61	VTTTSGYQCFSRYPDHMKQHDFPKSAMPE	GYVQERTIFFKDGNKYTKRAEVKEFGDTLV 120	
QY	121	NRIELKGIDFKEDGNILGHKLEYNNSHNV	YIMADKQNKGIANKFIRHNIEDGSVOLAD 180	
DB	121	NRIELKGIDFKEDGNILGHKLEYNNSHNV	YIMADKQNKGIANKFIRHNIEDGSVOLAD 180	
QY	181	HYOQNTPIGDGPVLLPDNNHYLSQSALS	KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238	
DB	181	HYOQNTPIGDGPVLLPDNNHYLSQSALS	KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238	
		RESULT 6		
		US-09-554-000-2		
		; Sequence 2, Application US/09554000		
		; Patent No. US20020165364A1		
		; GENERAL INFORMATION:		
		; APPLICANT: Tsielen, Roger Y.		
		; APPLICANT: Miyawaki, Atsushi		
		; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR		
		; TITLE OF INVENTION: DETECTION OF ANALYTES		
		; FILE REFERENCE: 07257/042001		
		; CURRENT APPLICATION NUMBER: US/09/554,000		
		; PRIOR FILING DATE: 2000-04-20		
		; PRIOR APPLICATION NUMBER: 08/818,252		
		; PRIOR FILING DATE: 1997-03-14		
		; NUMBER OF SEQ ID NOS: 56		
		; SOFTWARE: FastSeq for Windows Version 4.0		
		; SEQ ID NO 2		
		; LENGTH: 642		
		; TYPE: PRT		
		; ORGANISM: Aequorea victoria		
		US-09-554-000-2		
		Query Match	49.2%; Score 1259.5; DB 9; Length 642;	
		Best Local Similarity	57.5%; Pred. No. 3.3e-80;	
		Matches 263; Conservative	29; Mismatches 78; Indels 87; Gaps 5;	
QY	1	MSKGEELFTGVVPILVELDGNVGHKFSVS	GEGDATYGKLTFLKICITGKLPVPWPTL 60	
DB	2	VSKGEELFTGVVPILVELDGNVGHKFSVS	GEGDATYGKLTFLKICITGKLPVPWPTL 61	
QY	61	VTTLTGYQCFSRYPDHMKQHDFPKSAMPE	GYVQERTIFFKDGNKYTKRAEVKEFGDTLV 120	
DB	62	VTTLTGHVCFSRYPDHMKQHDFPKSAMPE	GYVQERTIFFKDGNKYTKRAEVKEFGDTLV 121	
		RESULT 7		
		US-09-554-000-4		
		; Sequence 4, Application US/09554000		
		; Patent No. US20020165364A1		
		; GENERAL INFORMATION:		
		; APPLICANT: Tsielen, Roger Y.		
		; APPLICANT: Miyawaki, Atsushi		
		; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR		
		; TITLE OF INVENTION: DETECTION OF ANALYTES		
		; FILE REFERENCE: 07257/042001		
		; CURRENT APPLICATION NUMBER: US/09/554,000		
		; CURRENT FILING DATE: 2000-04-20		
		; PRIOR APPLICATION NUMBER: 08/818,252		
		; PRIOR FILING DATE: 1997-03-14		
		; NUMBER OF SEQ ID NOS: 56		
		; SOFTWARE: FastSeq for Windows Version 4.0		
		; SEQ ID NO 4		
		; LENGTH: 652		
		; TYPE: PRT		
		; ORGANISM: Aequorea victoria		
		US-09-554-000-4		
		Query Match	49.2%; Score 1259.5; DB 9; Length 652;	
		Best Local Similarity	57.5%; Pred. No. 3.3e-80;	
		Matches 263; Conservative	29; Mismatches 78; Indels 87; Gaps	
QY	1	MSKGEELFTGVVPILVELDGNVGHKFSVS	GEGDATYGKLTFLKICITGKLPVPWPTL 60	
DB	2	VSKGEELFTGVVPILVELDGNVGHKFSVS	GEGDATYGKLTFLKICITGKLPVPWPTL 61	
QY	61	VTTLTGYQCFSRYPDHMKQHDFPKSAMPE	GYVQERTIFFKDGNKYTKRAEVKEFGDTLV 120	
DB	62	VTTLTGHVCFSRYPDHMKQHDFPKSAMPE	GYVQERTIFFKDGNKYTKRAEVKEFGDTLV 121	
QY	121	NRIELKGIDFKEDGNILGHKLEYNNSHNV	YIMADKQNKGIANKFIRHNIEDGSVOLAD 180	
DB	122	NRIELKGIDFKEDGNILGHKLEYNNSHNV	YIMADKQNKGIANKFIRHNIEDGSVOLAD 181	
QY	181	HYOQNTPIGDGPVLLPDNNHYLSQSALS	KDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240	
DB	182	HYOQNTPIGDGPVLLPDNNHYLSQSALS	KDPNEKRDMHVLLEFVTAARHMDOLTE----- 236	
QY	241	GSGGSGGGSGGGSGGGSGGGSGGGSGGG	SGGGSGGLSVKLTSDFDNPWR 300	
DB	237	-----	-----EQ 238	
QY	301	IGRHKHMFNFLDVNHHNGKISLDEMYYKAS	DIVINNLGATPQAKRKHKDAVEAFFCGAGMK 360	
DB	239	IAEFKEAFSLFDKDGDTITTTEL-----	GTVMRS	

```

Qy 361 YGVETDWPAYIE--GWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKDQQAITLDWK 418
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 Y-----FPEFTLMARKMKTDSEIEI-----AFRVFDKDGNGYISAAELR 336
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 419 AYTKAAGTIQSSDCEETFRVCDIDESQLDVDEMTR 455
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 337 HVMNGLKEKLTDEEVDEMIREADIDGGQVNYEEFVQ 373
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-09-900-345A-60
; Sequence 60, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; TITLE OF INVENTION: EFFICIENCY OF A CODON
; FILE REFERENCE: 10338-505
; CURRENT APPLICATION NUMBER: US/09/900,345A
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 60
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTA)5GFP
US-09-900-345A-60

Query Match          49.2%; Score 1259; DB 9; Length 243;
Best Local Similarity 98.3%; Pred. No. 1.le-80;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSGEELFTGVPIVLVDGDNVNGHKFVSVEGEDGATYGLTLKFCITCTGKLPVWPPTL 60
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6 LSKGEELFTGVPIVLVDGDNVNGHKFVSVEGEDGATYGLTLKFCITCTGKLPVWPPTL 65
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 61 VTTLTVGVQCFSRYPDHMKQHDFFKAMPYGYVQERTIFFKDDGNGYKTRAEVFEGETLV 120
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 66 VTTFSGVQCFSRYPDHMKQHDFFKAMPYGYVQERTIFFKDDGNGYKTRAEVFEGETLV 125
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 121 NRTELKGDIFKEDGNILGHLEKLYNSHNHYINADKQKIGKANFKIRHNIEDGSVOLAD 180
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 126 NRTELKGDIFKEDGNILGHLEKLYNSHNHYINADKQKIGKANFKIRHNIEDGSVOLAD 185
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGTTHGMDELYK 238
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGTTHGMDELYK 243
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-900-345A-62
; Sequence 62, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; TITLE OF INVENTION: EFFICIENCY OF A CODON
; FILE REFERENCE: 10338-505
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07

```



```

QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 238
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 243

RESULT 11
US-09-900-345A-66
; Sequence 66, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; TITLE OF INVENTION: EFFICIENCY OF A CODON
; FILE REFERENCE: 10338-5US
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/000008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTT)5GFP
US-09-900-345A-66

Query Match 49.2%; Score 1259; DB 9; Length 243;
Best Local Similarity 98.3%; Pred. No. 1.1e-80;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTFLKFICTTGKLPVPWPTL 60
DB 6 LSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTFLKFICTTGKLPVPWPTL 65

QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTTFSGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTLV 125

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKANFKIRHINIEDGSSVOLAD 180
DB 126 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKANFKIRHINIEDGSSVOLAD 185

QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 238
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 243

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTT)5GFP
US-09-900-345A-66

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(TTA)5GFP
US-09-900-345A-68

Query Match 49.2%; Score 1259; DB 9; Length 243;
Best Local Similarity 98.3%; Pred. No. 1.1e-80;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTFLKFICTTGKLPVPWPTL 60
DB 6 LSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTFLKFICTTGKLPVPWPTL 65

QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTTFSGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTLV 125

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKANFKIRHINIEDGSSVOLAD 180
DB 126 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKANFKIRHINIEDGSSVOLAD 185

QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 238
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 243

RESULT 13
US-09-900-345A-70
; Sequence 70, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; TITLE OF INVENTION: EFFICIENCY OF A CODON
; FILE REFERENCE: 10338-5US
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/000008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(TTG)5GFP
US-09-900-345A-70

```

```

RESULT 12
US-09-900-345A-68
; Sequence 68, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; TITLE OF INVENTION: EFFICIENCY OF A CODON
; FILE REFERENCE: 10338-5US
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/000008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 68
; LENGTH: 243
; TYPE: PRT

```

```

Query Match 49.2%; Score 1259; DB 9; Length 243;
Best Local Similarity 98.3%; Pred. No. 1.1e-80;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTFLKFICTTGKLPVPWPTL 60
DB 6 LSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGKLTFLKFICTTGKLPVPWPTL 65

QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTTFSGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKFEGDTLV 125

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKANFKIRHINIEDGSSVOLAD 180
DB 126 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQNGIKANFKIRHINIEDGSSVOLAD 185

QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 238
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK 243

```

## RESULT 14

US-10-057-505-2  
; Sequence 2, Application US/10057505  
; Patent No. US20020164674A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: AURORA BIOSCIENCES CORPORATION  
; APPLICANT: TSIEN, Roger  
; APPLICANT: HEIM, Roger  
; APPLICANT: CUBITT, Andrew  
; TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS  
; FILE REFERENCE: REG1260-3  
; CURRENT APPLICATION NUMBER: US/10/057,505  
; CURRENT FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: US 08/792,553  
; PRIOR FILING DATE: 1997-01-31  
; PRIOR APPLICATION NUMBER: US 09/396,003  
; PRIOR FILING DATE: 1999-09-13  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Aequorea victoria  
US-10-057-505-2

Query Match 49.1%; Score 1258; DB 9; Length 238;  
Best Local Similarity 98.3%; Pred. No. 1.3e-80;  
Matches 234; Conservative 2; Mismatches 2; Indels 0; Caps 0;  
QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFICTTGKLPVWPPTL 60  
DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFICTTGKLPVWPPTL 60  
QY 61 VTLLTYGQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120  
DB 61 VTTSYGVQCFSRYPDHMKRHDFKFSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120  
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180  
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180  
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

## RESULT 15

US-09-884-681-2  
; Sequence 2, Application US/09884681  
; Patent No. US20020061546A1  
; GENERAL INFORMATION:  
; APPLICANT: Tsiien, Roger Y.  
; APPLICANT: Cubitt, Andrew B.  
; TITLE OF INVENTION: Assays for Protein Kinases Using  
; Fluorescent Protein Substrates  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/884,681  
; FILING DATE: 19-Jun-2001  
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/679,865  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Storella, John S.  
; REGISTRATION NUMBER: 32,944  
; REFERENCE/DOCKET NUMBER: 023072-069000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 238 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-884-681-2

Query Match 49.1%; Score 1258; DB 10; Length 238;  
Best Local Similarity 98.3%; Pred. No. 1.3e-80;  
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFICTTGKLPVWPPTL 60  
DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFICTTGKLPVWPPTL 60  
QY 61 VTLLTYGQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120  
DB 61 VTTSYGVQCFSRYPDHMKRHDFKFSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120  
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180  
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180  
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

Search completed: March 13, 2003, 15:12:35  
Job time : 23.9127 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:01:59 ; Search time 27.1537 Seconds  
(without alignments)  
1688.761 Million cell updates/sec

Title: US-09-863-901-5  
Perfect score: 2560  
Sequence: 1 MSKGELFTGVVPILVELDG.....LGFWYTMDSACEKLYGGAVP 477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1256	49.1	238	1 JQ1514	green-fluorescent
2	1032	40.3	196	1 AQJENV	aequorin precursor
3	962	37.6	196	2 A26623	aequorin-l precursor
4	742	29.0	198	2 S39022	mitrocomin precursor
5	663	25.9	198	2 S28860	claytin - hydromedu
6	170	6.6	302	2 C84470	hypothetical prote
7	168	6.6	592	2 E82759	endo-1,4-beta-gluc
8	161	6.3	115	2 D61615	sericin MG-1 - gre
9	159.5	6.2	207	2 T07381	glycine-rich prote
10	158.5	6.2	268	1 C1HUL	calpain (EC 3.4.22
11	155.5	6.1	1226	2 T24045	hypothetical prote
12	155	6.1	641	1 Q0BE31	nuclear antigen EB
13	155	6.1	1218	2 E84537	hypothetical prote
14	154.5	6.0	622	2 I37984	keratin 9, type I,
15	154	6.0	1585	2 T31611	hypothetical prote
16	151.5	5.9	481	2 A35628	loricrin - mouse
17	149.5	5.8	204	2 T09592	protein corA, cold
18	149.5	5.8	221	2 T04592	glycine-rich cell
19	148.5	5.8	171	2 H84709	probable glycine-r
20	148	5.8	291	1 S31415	glycine-rich prote
21	148	5.8	1275	2 T49362	hypothetical prote
22	148	5.8	1901	2 F70806	hypothetical glyci
23	147.5	5.8	165	1 KNRZG1	glycine-rich cell
24	147.5	5.8	434	1 Z3BPIK	coat protein A - p
25	147	5.7	183	2 PNO109	keratin-like prote
26	147	5.7	526	1 K8EOVI	keratin, 54K type
27	147	5.7	569	1 KRMSEI	keratin, 59K type
28	147	5.7	995	2 T22942	hypothetical prote
29	146.5	5.7	266	1 CIRBL	calpain (EC 3.4.22

30	146	5.7	465	1 S01820	glycine-rich cell
31	145.5	5.7	166	1 KRBO2B	keratin, 68K type
32	145	5.7	263	2 A34466	calpain (EC 3.4.22
33	145	5.7	320	2 T09555	fibrillarlin - Arab
34	145	5.7	988	2 T08102	myrosinase-binding
35	144.5	5.6	266	1 C1PGL	calpain (EC 3.4.22
36	144.5	5.6	462	4 S33798	FUS/CHOP mutant fu
37	144	5.6	183	1 KNRZG2	glycine-rich cell
38	143.5	5.6	208	2 T46896	merozoite surface
39	143.5	5.6	434	2 S08091	gene III protein -
40	143.5	5.6	593	1 KRHU0	keratin 10, type I
41	143	5.6	150	2 C86224	hypothetical prote
42	143	5.6	420	2 A49642	transcription fact
43	142.5	5.6	167	2 S21359	keratin, type I, c
44	142.5	5.6	271	2 S34666	glycine-rich prote
45	142.5	5.6	528	2 G02127	fus-like protein -

ALIGNMENTS

RESULT 1

JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 23-Mar-2001

C:Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331

R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.

Gene 111, 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: JQ1514; MUID:9217527; PMID:1347277

A:Accession: JS0692

A:Molecule type: DNA

A:Residues: 1-107 'S', 109-238 <PRA1>

A:Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663

A:Accession: JQ1514

A:Molecule type: mRNA

A:Residues: 1-99 'F', 101-140 'L', 142-218 'V', 220-238 <PRA2>

A:Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661

A:Accession: PQ0335

A:Molecule type: protein

A:Residues: 46-64,74-122,132-151;154-183;185-200 <PRA3>

R:Inouye, S.; Tsuji, F.I.

FEBS Lett. 351, 211-214, 1994

A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-24 'Q', 26-156 'P', 158-171 'K', 173-238 <INO>

A:Cross-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384

R:Watkins, J.N.; Campbell, A.K.

submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Accession: S51330

A:Molecule type: mRNA

A:Residues: 1-13 'V', 15-24 'Q', 26-44 'N', 46-153 'G', 155-156 'P', 158-171 'K', 173-227,

A:Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009

A:Experimental source: Clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24 'Q', 26-29 'R', 31-83 'L', 85-153 'G', 155-156 'P', 158-171 'K', 173-208,

A:Cross-references: EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PID:g634011

A:Experimental source: Clone gfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 8

A>Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Not. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

A:Contents: annotation; X-ray crystallography, 1.9 angstroms  
 C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting  
 C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
 C:Genetics: GFP  
 A:Gene: GFP  
 A:Introns: 69/3; 167/3  
 C:Superfamily: green-fluorescent protein  
 C:Keywords: chromoprotein; luminescence  
 F:65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental  
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 49.1%; Score 1256; DB 1; Length 238;  
 Best Local Similarity 97.5%; Pred. No. 3.7e-73;  
 Matches 232; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSAGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTILKICTCKLVPWPTL 60  
 DB 1 MSAGEELFTGVVPILVELDGDVNGHKFVSQGEKGDATYKLTILKICTCKLVPWPTL 60  
 QY 61 VTTLTGVGOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120  
 DB 61 VTTFSGVGOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120  
 QY 121 NRIELKGIDKEDNGILGKLEYNNSHVYIMADKOKNGTKANFKIRHNIEDGSVQLAD 180  
 DB 121 NRIELKGIDKEDNGILGKMEYNNSHVYIMADKOKNGIKVFNKIRHNIEDGSVQLAD 180  
 QY 181 HYQONTPTGDGVLPDPNHYLSTQSLSKDPNEKRDRHMLLEFVTAAGITGHGMDLYK 238  
 DB 181 HYQONTPTGDGVLPDPNHYLSTQSLSKDPNEKRDRHMLLEFVTAAGITGHGMDLYK 238

## RESULT 2

AQJFNV  
 aequorin precursor - hydromedusa (Aequorea victoria)  
 C:Species: Aequorea victoria  
 C:Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 26-May-2000  
 C:Accession: A03020  
 R:Inouye, S.; Noguchi, M.; Sakaki, Y.; Takagi, Y.; Miyata, T.; Iwanaga, S.; Miyata, T.;  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3154-3158, 1985  
 A:Title: Cloning and sequence analysis of cDNA for the luminescent protein aequorin.  
 A:Reference number: A03020; MUID:85216460; PMID:3858813  
 A:Accession: A03020  
 A:Molecule type: mRNA  
 A:Residues: 1-196 <INO>  
 A:Cross-references: GB:M11394; NID:g155658; PIDN:AAA27719.1; PID:g155659; GB:L29571; NID  
 A:Experimental source: clone AQ440  
 C:Comment: The precise function of residues 1-7 is not known.  
 C:Comment: The authors suggest that there are three calcium-binding sites, and that resi  
 C:Comment: Trace amounts of calcium ion trigger the oxidation of the functional chromoph  
 C:Comment: Cysteine residues appear to be critical to the ability of this photoprotein t  
 n, and mercaptoethanol is necessary to regenerate aequorin from apoaequorin.  
 C:Superfamily: calmodulin; calmodulin repeat homology  
 F:8-196/Product: aequorin #status predicted <E1>  
 F:11-143/Domain: calmodulin repeat homology <E1>  
 F:147-179/Domain: calmodulin repeat homology <E2>  
 F:31,33,37,42/Binding site: calcium (Asp, Asn, Lys, Glu) #status predicted  
 F:124,126,128,130,135/Binding site: calcium (Asp, Asp, Asn, Ala, Glu) #status predicted  
 F:160,162,164,166,171/Binding site: calcium (Asp, Asp, Ser, Glu, Glu) #status predicted

Query Match 40.3%; Score 1032; DB 1; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-59;  
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWIGRHKHMFNFDVNHGKISLDEVMYKASDIVINNLGATPEQAKRHK 347  
 DB 7 SVKLTSDFDNPRWIGRHKHMFNFDVNHGKISLDEVMYKASDIVINNLGATPEQAKRHK 66  
 QY 348 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407

Db 67 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 126  
 QY 408 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467  
 Db 127 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 186  
 QY 468 CEKLYGGAVP 477  
 Db 187 CEKLYGGAVP 196

## RESULT 3

A26623  
 aequorin-l precursor - hydromedusa (Aequorea victoria)  
 C:Species: Aequorea victoria  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-May-2000  
 C:Accession: A26623  
 R:Prasher, D.C.; McCann, R.O.; Longiaru, M.; Cormier, M.J.  
 Biochemistry 26, 1326-1332, 1987  
 A:Title: Sequence comparisons of complementary DNAs encoding aequorin isoforms.  
 A:Reference number: A26623; MUID:87185437; PMID:2882777  
 A:Accession: A26623  
 A:Molecule type: mRNA  
 A:Residues: 1-196 <PRA>  
 A:Cross-references: GB:M16103; NID:g155652; PIDN:AAA27716.1; PID:g155653  
 A:Note: the authors translated the codon GAT for residue 143 as Ala  
 C:Comment: The precise function of residue 1-7 is not known.  
 C:Superfamily: calmodulin; calmodulin repeat homology  
 C:Keywords: calcium binding; EF hand; luminescence  
 F:18-50/Domain: calmodulin repeat homology <E1>  
 F:111-143/Domain: calmodulin repeat homology <E2>  
 F:147-179/Domain: calmodulin repeat homology <E3>

Query Match 37.6%; Score 962; DB 2; Length 196;  
 Best Local Similarity 90.5%; Pred. No. 1.7e-54;  
 Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWIGRHKHMFNFDVNHGKISLDEVMYKASDIVINNLGATPEQAKRHK 347  
 Db 7 SVKLTSDFDNPRWIGRHKHMFNFDVNHGKISLDEVMYKASDIVINNLGATPEQAKRHK 66  
 QY 348 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407  
 Db 67 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 126  
 QY 408 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467  
 Db 127 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 186  
 QY 468 CEKLYGGAVP 477  
 Db 187 CEKLYGGAVP 196

## RESULT 4

S39022  
 Mitrocomin precursor - Mitrocoma cellularia  
 C:Species: Mitrocoma cellularia  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 26-May-2000  
 C:Accession: S39022  
 R:Fagan, T.F.; Ohmura, Y.; Blinks, J.R.; Inouye, S.; Tsuji, F.I.  
 FEBS Lett. 333, 301-305, 1993  
 A:Title: Cloning, expression and sequence analysis of cDNA for the Ca(2+)-binding pho  
 A:Reference number: S39022; MUID:94039781; PMID:8224198  
 A:Accession: S39022  
 A:Molecule type: mRNA  
 A:Residues: 1-198 <FAG>  
 A:Cross-references: EMBL:L31623; NID:9468907; PIDN:AAA29298.1; PID:g468908  
 C:Superfamily: calmodulin; calmodulin repeat homology  
 C:Keywords: calcium binding; EF hand; luminescence  
 F:19-51/Domain: calmodulin repeat homology <E1>  
 F:112-144/Domain: calmodulin repeat homology <E2>  
 F:148-180/Domain: calmodulin repeat homology <E3>

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C

R;Zurovec, M.; Sehnal, F.; Scheller, K.; Kumaran, A.K.





Db 489 RGGSGSGYGGGSGGSGRGGSGGSGYGGGSGGGSG 530

RESULT 15

T31611

hypothetical protein Y50E8A.g - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T31611

R:Steward, C.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z21047

A:Accession: T31611

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1585 <WIL>

A:Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB5050.1; CESP:Y50E8A.g

A:Experimental source: clone Y50E8A

C:Genetics:

A:Gene: CESP:Y50E8A.g

A:Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Query Match 6.0%; Score 154; DB 2; Length 1585;

Best Local Similarity 23.3%; Pred. No. 0.072;

Matches 51; Conservative 26; Mismatches 46; Indels 96; Gaps 8;

QY 131 KEDGNLGHKLEYNHNVIMADKQNGIKANFKIRHNIEDGVSQVLAHYOQNTPIGD 190

Db 660 EEDTQLPHRRHHHHHH-----OLQLQHHLED-----T 691

QY 191 GPVLLPDNHYLSTQSALSKDPNEKRDMVLEFVTA-----GITHGMDL----- 236

Db 692 HPEEVEDRH-----EVVMLQEDDAEDDLVVVEEDTQLPHRRHHHQLQHHLEDTHPEST 747

QY 237 -----YKSGGSGSGSGSGSG-----SGG 255

Db 748 IWRIRIRGGSGSGGYASGGGGSGSGGSGGGRYSSAAAPPPPPPPAPAPAPSSGG 807

QY 256 QSGS-----GSGGSGSGSGSGSGSG--SGGSG 285

Db 808 YSGSGSDSAGGGGSGSRGYSEGGSSGGGGSGGYSG 846

Search completed: March 13, 2003, 15:05:45

Job time : 31.1537 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:00:34 ; Search time 15.3871 Seconds  
(without alignments)  
1285.766 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGEELFTGVVPILVELDG.....LGFWYTMPACERLYGGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	1262	49.3	238	1	GFP_AEQVI	P42212 aequorea vi
2	1032	40.3	196	1	AEQ2_AEQVI	P02592 aequorea vi
3	962	37.6	196	1	AEQ1_AEQVI	P07164 aequorea vi
4	742	29.0	198	1	MYTR_MITCE	P39047 mitrocoma c
5	718	28.0	195	1	OBL_OBELA	Q27709 obelia long
6	663	25.9	198	1	CLYT_CLYGR	Q08121 clytia greg
7	171.5	6.7	532	1	2IC2_HUMAN	O95409 homo sapien
8	161.5	6.3	450	1	SWP1_ENCCU	Q9XZV1 encephalito
9	161	6.3	115	1	SER1_GALME	O96614 gallieria me
10	158.5	6.2	268	1	CANS_HUMAN	P04632 homo sapien
11	155	6.1	641	1	EBN1_EBV	P03211 Epstein-bar
12	154.5	6.0	622	1	K1C1_HUMAN	P35527 homo sapien
13	151.5	5.9	481	1	LORI_MOUSE	P18165 mus musculu
14	149.5	5.8	204	1	CORA_MEDSA	Q07202 medicago sa
15	149	5.8	269	1	CANS_MOUSE	O88456 mus musculu
16	149	5.8	421	1	BR3A_MOUSE	P17208 mus musculu
17	148.5	5.8	627	1	K2C1_MOUSE	O70133 mus musculu
18	148	5.8	1380	1	DDX9_MOUSE	O53553 mycobacteri
19	148	5.8	1901	1	Y208_MYCTU	P25074 oryza sativ
20	147.5	5.8	165	1	GRP1_ORYSA	P03663 bacterioph
21	147.5	5.8	434	1	COAA_BPKE	P06394 bos taurus
22	147	5.7	526	1	K1CJ_BOVIN	P02535 mus musculu
23	147	5.7	569	1	K1CJ_MOUSE	P06813 oryctolagus
24	146.5	5.7	266	1	CANS_RABIT	P10496 phaseolus v
25	146	5.7	465	1	GRP2_PHAVU	P04262 bos taurus
26	145.5	5.7	166	1	K2C5_BOVIN	P13135 bos taurus
27	145	5.7	263	1	CANS_BOVIN	P04574 sus scrofa
28	144.5	5.6	266	1	CANS_PIG	P29834 oryza sativ
29	144	5.6	183	1	GRP2_ORYSA	Q64537 rattus norv
30	143.5	5.6	266	1	CANS_RAT	P15415 bacterioph
31	143.5	5.6	434	1	COAA_BPT22	P13645 homo sapien
32	143.5	5.6	593	1	K1CJ_HUMAN	P34643 dictyostel
33	143.5	5.6	677	1	SPR7_DICDI	

## RESULT 1

ID	GFP_AEQVI	STANDARD	PRT	238 AA
AC	P42212: Q17104;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	(1)			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent protein.";			
RL	Gene 111:229-233(1992).			
RN	(2)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
RN	(3)			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein.";			
RL	Biochemistry 32:1212-1218(1993).			
RN	(4)			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormoie M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RA	Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent protein.";			
RL	Science 273:1392-1395(1996).			
RN	(5)			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			
RN	(6)			
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.			
RX	MEDLINE=98455509; PubMed=9782051;			
RA	Wachter R.M., Ellsiger M.A., Kallio K., Hanson G.T., Remington S.J.;			
RT	"Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein.";			
RL	Structure 6:1267-1277(1998).			

## ALIGNMENTS

34	141.5	5.5	759	1	GSPD_XANCP	P29041 xanthomonas
35	141.5	5.5	1250	1	TP3A_DROME	Q9ng98 drosophila
36	141	5.5	1726	1	MSP1_PLAFC	P04934 plasmodium
37	141	5.5	1726	1	MSP1_PLAFC	P50495 plasmodium
38	140	5.5	322	1	PUR_HUMAN	Q00577 homo sapien
39	140	5.5	333	1	SIX3_MOUSE	Q62233 mus musculu
40	139.5	5.4	316	1	LORI_HUMAN	P23490 homo sapien
41	139.5	5.4	384	1	GRP1_PETHY	P09789 petunia hyb
42	139	5.4	424	1	COAA_BPFD	P03661 bacterioph
43	139	5.4	424	1	COAA_BPFD	P03662 bacterioph
44	138.5	5.4	148	1	CLM4_MOUSE	Q9jms3 mus musculu
45	138.5	5.4	734	1	YKR2_CABEL	P34308 caenorhabdi

```

[7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RA MEDLINE=93238303; PubMed=10220315;
RA Elsiger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
RT "Structural and spectral response of green fluorescent protein
RT variants to changes in pH."
RL Biochemistry 38:5296-5301(1999).
CC -!- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE
CC BLUE CHEMILUMINESCENCE OF THE PROTEIN Aequorin INTO GREEN
CC FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON
CC RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN Aequorin.
CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
CC ABSORBANCE PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: PHOTOCYTES.
CC -!- PTM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.
CC -!- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
CC chimeric proteins of GFP linked to other proteins where it
CC functions as a fluorescent protein tag. GFP tolerates N- and C-
CC terminal fusion to a broad variety of proteins. It has been
CC expressed in bacteria, yeast, slime mold, plants, drosophila,
CC zebrafish, and in mammalian cells. As a noninvasive fluorescent
CC marker in living cells, it allows for a wide range of applications
CC where it may function as a cell lineage tracer, reporter of gene
CC expression, or as a measure of protein-protein interactions.
CC -!- DATABASE: NAME-Protein Spotlight;
CC NOTE-Issue 11 of June 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sptl011.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M62654; AAA27722.1; -
DR EMBL: M62653; AAA27721.1; -
DR EMBL: L29345; AAA58246.1; -
DR PDB: 1GFL; 11-JAN-97.
DR PDB: 1EMA; 08-NOV-96.
DR PDB: 1EMB; 16-JUN-97.
DR PDB: 1EMC; 20-AUG-97.
DR PDB: 1EMG; 20-AUG-97.
DR PDB: 2EMD; 20-AUG-97.
DR PDB: 1EME; 20-AUG-97.
DR PDB: 1EMF; 20-AUG-97.
DR PDB: 1EMG; 12-MAY-99.
DR PDB: 1EMK; 20-AUG-97.
DR PDB: 1EMI; 20-AUG-97.
DR PDB: 1EMW; 20-AUG-97.
DR PDB: 2EMN; 20-AUG-97.
DR PDB: 2EMO; 20-AUG-97.
DR PDB: 1BFP; 07-JUL-97.
DR PDB: 1YFP; 28-OCT-98.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFPLORESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
KW Luminescence; 3D-structure.
FT SITE 65 67 MODIFIED TO FORM THE CHROMOPHORE.
FT VARIANT 100 100 F -> Y.
FT VARIANT 108 108 T -> S.
FT VARIANT 141 141 L -> M.
FT VARIANT 219 219 V -> I.
FT CONFLICT 25 25 H -> Q (IN REF. 2).
FT CONFLICT 157 157 Q -> P (IN REF. 2).
FT CONFLICT 172 172 E -> K (IN REF. 2).
FT SEQUENCE 238 AA; 26886 MW; EA5A6F21F6FB6E05 CRC64;

```

```

Query Match 49.3%; Score 1262; DB 1; Length 238;
Best Local Similarity 98.7%; Pred. No. 3e-69;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTLLKFKICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGDGYGKLTLLKFKICTTGKLPVWPPTL 60
QY 61 VTTTLYGQVCSFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEFEGDTLV 120
DB 61 VTTTLYGQVCSFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLLEYNHNHVMADKOKNGIKANFKIRHNIEDGSVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLLEYNHNHVMADKOKNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQNTPTIGDGPVLLPDNHYLSTQSLSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQNTPTIGDGPVLLPDNHYLSTQSLSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 2
AEQ2_AEQVI STANDARD; PRT; 196 AA.
ID AEQ2_AEQVI
AC P02592;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aequorin 2 precursor.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RX SEQUENCE FROM N.A.
RX MEDLINE=85216460; PubMed=3858813;
RA Inouye S., Noguchi M., Sakaki Y., Takagi Y., Miyata T., Iwanaga S.,
RA Miyata T., Tsuji F.I.;
RT "Cloning and sequence analysis of cDNA for the luminescent protein
RT aequorin."
RT proc. Natl. Acad. Sci. U.S.A. 82:3154-3158(1985).
[2]
RX SEQUENCE OF 9-185 FROM N.A. (AEQUORIN 2 AND 3).
RX MEDLINE=87185437; PubMed=2882777;
RA Prasher D.C., McCann R.O., Longiaru M., Cormier M.J.;
RT "Sequence comparisons of complementary DNAs encoding aequorin
RT isotypes."
RT Biochemistry 26:1326-1332(1987).
[3]
RX SEQUENCE OF 8-196.
RX MEDLINE=86077721; PubMed=2866797;
RA Charbonneau H., Walsh K.A., McCann R.O., Prendergast F.G.,
RA Cormier M.J., Vanaman T.C.;
RT "Amino acid sequence of the calcium-dependent photoprotein aequorin."
RT Biochemistry 24:6762-6771(1985).
[4]
RX MUTAGENESIS.
RA Tsuji F.I., Inouye S., Goto T., Sakaki Y.;
RT "Site-specific mutagenesis of the calcium-binding photoprotein
RT aequorin."
RL proc. Natl. Acad. Sci. U.S.A. 83:8107-8111(1986).
[5]
RX MUTAGENESIS OF PRO-196.
RX MEDLINE=92111761; PubMed=1765170;
RA Nomura M., Inouye S., Ohmiya Y., Tsuji F.I.;
RT "A C-terminal proline is required for bioluminescence of the Ca(2+)-
RT binding photoprotein, aequorin."
RL FEBS Lett. 295:63-66(1991).
[6]
RX DISULFIDE BOND.
RA Ohmiya Y., Kuroso S., Ohashi M., Fagan T.F., Tsuji F.I.;
RT "Mass spectrometric evidence for a disulfide bond in aequorin

```



```

DR Pfam: PF000036; ehand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF_HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 7
FT CHAIN 8 196
FT SITE 47 57
FT SITE 62 72
FT SITE 107 117
FT CA_BIND 31 42
FT DOMAIN 72 88
FT CA_BIND 124 135
FT CA_BIND 160 171
FT DISULFID 152 159
SQ SEQUENCE 196 AA; 22514 MW; 9AA5B636288A5B8F CRC64;

Query Match 37.6%; Score 962; DB 1; Length 196;
Best Local Similarity 90.5%; Pred.No. 2.3e-51;
Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 288 SVKLTDFDNPRTWGRHKHMFNFDVNHNGKISLDEMYKASDIVNNLGATPEQAKRHK 347
DB 7 SVKLTDFDNPRTWGRHKHMFNFDVNHNGKISLDEMYKASDIVNNLGATPEQAKRHK 56
QY 348 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKD 407
DB 67 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKD 126
QY 408 QNGAITLDWKATKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMOPA 467
DB 127 QNGAISLDEWKATKSDGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMOPA 186
QY 468 CEKLYGGAVP 477
DB 187 CEKLYGGAVP 196

RESULT 4
MYTR_MITCE
ID MYTR_MITCE STANDARD; PRT; 198 AA.
AC F39047;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitrocomin precursor.
GN M17.
OS Mitrocoma cellularia (Halistaura mitrocoma).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;
OC Mitrocomidae; Mitrocoma.
OX NCBI_TaxID=31874;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039781; PubMed=8224198;
RA Fagan T.F., Ohmura Y., Blinks J.R., Inouye S., Tsuji F.I.;
RT "Cloning, expression and sequence analysis of cDNA for the Ca(2+)-
RT binding photoprotein, mitrocomin."
RL FEBS Lett. 333:301-305(1993).
CC -1- FUNCTION: CA(2+)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
CC AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
CC ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
CC COLENTERAZINE INTO COLENTERAMIDE AND CO(2) WITH THE
CC CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. Aequorin family.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: L31623; AAA29298.1; -.
DR PIR: S39022; S39022.
DR HSSP: P02592; 1EJ3.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF_HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 8
FT CHAIN 9 198
FT CA_BIND 32 43
FT DOMAIN 73 89
FT CA_BIND 125 136
FT CA_BIND 161 172
FT DISULFID 153 160
SQ SEQUENCE 198 AA; 22714 MW; 8F6307EF0966F670 CRC64;

Query Match 29.0%; Score 742; DB 1; Length 198;
Best Local Similarity 67.9%; Pred.No. 3.6e-38;
Matches 129; Conservative 29; Mismatches 32; Indels 0; Gaps 0;

QY 288 SVKLTDFDNPRTWGRHKHMFNFDVNHNGKISLDEMYKASDIVNNLGATPEQAKRHK 347
DB 8 AVKLTDFDNPRTWGRHKHMFNFDVNHNGKISLDEMYKASDIVNNLGATPEQAKRHK 67
QY 348 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKD 407
DB 68 KCVDFGGAGLEYDQDTPTEYIEGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKD 127
QY 408 QNGAITLDWKATKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMOPA 467
DB 128 RGVSLDEWQYTHCAGIQSRGQCEATFAHCDLDGDKLDVDEMTROHLGFWYTMOPA 187
QY 468 CEKLYGGAVP 477
DB 188 CEKLYGGAVP 197

RESULT 5
OBL_OBELO
ID OBL_OBELO STANDARD; PRT; 195 AA.
AC Q27709;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Obelin precursor (OBL).
OS Obelia longissima (Black sea hydrozoan).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Campanulariidae; Obelia.
OX NCBI_TaxID=32570;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95180731; PubMed=7875600;
RA Iliarionov B.A., Bondar V.S., Iliarionova V.A., Vysotski E.S.;
RT "Sequence of the cDNA encoding the Ca(2+)-activated photoprotein
RT obelin from the hydroid polyp Obelia longissima."
RL Gene 153:273-274(1995).
CC -1- FUNCTION: CA(2+)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
CC AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
CC ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
CC COLENTERAZINE INTO COLENTERAMIDE AND CO(2) WITH THE
CC CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. Aequorin family.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC EMBL: U07128; AAA67708.1; -.
CC HSSP: P02592; 1EJ3.
CC InterPro: IPR002048; EF-hand.
CC Pfam: PF00036; efhand; 3.
CC ProDom: PD000012; EF-hand; 1.
CC SMART: SM00054; EFh; 2.
CC PROSITE: PS00018; EF-hand; 3.
CC Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 195
FT CHAIN 7 195
FT CA_BIND 30 41
FT DOMAIN 71 87
FT CA_BIND 123 134
FT CA_BIND 159 170
FT CA_BIND 151 158
FT DISULFID 151 158
SQ SEQUENCE 195 AA; 22226 MW; 5D002270B73D3663 CRC64;

Query Match 28.0%; Score 718; DB 1; Length 195;
Best Local Similarity 68.4%; Pred. No. 9.8e-37;
Matches 130; Conservative 23; Mismatches 37; Indels 0; Gaps 0;

Qy 28 SVKLTSDFDNPRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347
Db 6 AVKLKTFDNDPRWIKRHKHMFDFLDINGNGKITLDEIVSKASDDICAKLEATPEQTKRHQ 65
Qy 348 DAVEAFFGAGMKYGVETDPAYTEGWKKLATDELYAKNEPTLIRIWGALFDIVDKD 407
Db 66 VCVEAFFGCGMEYKEITAFQFDGKQLATSELUKWARNEPTLIRIWGAQVDFIDKD 125
Qy 408 ONGATILDEWKAYTKAAGIIOSSEDCETFRVCDIDESSGOLDVDVDMTRHGLGFYWTMDPA 467
Db 126 GSGTITLDEWKAYKIGSISQEDCEATFRHCDLDSGDLVDVDMTRHGLGFYWTLDPE 185
Qy 468 CEKLYGGAVP 477
Db 186 ADGLYNGVP 195

RESULT 6
CLYT_CLYGR STANDARD; PRT; 198 AA.
AC Q08121;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clytin precursor (Phalidid).
OS Clytia gregaria (Phalidium gregarium).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Campanularidae; Clytia.
OX NCBI_TaxID=27801;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=93138101; PubMed=8422928;
RA Inouye S., Tsuji F.I.;
RT "Cloning and sequence analysis of cDNA for the Ca(2+)-activated
RT photoprotein, clytin."
RL FEBS Lett. 315:343-346(1993).
CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
CC AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
CC ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
CC COLENTERAZINE INTO COELENTERAMIDE AND CO(2) WITH THE
CC CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC EMBL: L13247; AAA28293.1; -.
CC EMBL: X70221; CAA49754.1; -.
CC PIR: S28860; S28860.
CC HSSP: P02592; 1EJ3.
CC InterPro: IPR002048; EF-hand.
CC Pfam: PF00036; efhand; 3.
CC ProDom: PD000012; EF-hand; 1.
CC SMART: SM00054; EFh; 2.
CC PROSITE: PS00018; EF-hand; 3.
CC Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 198
FT CHAIN 10 198
FT CA_BIND 33 44
FT DOMAIN 74 90
FT CA_BIND 126 137
FT CA_BIND 162 173
FT DISULFID 148 161
SQ SEQUENCE 198 AA; 22541 MW; 23F1E399667F9059 CRC64;

Query Match 25.9%; Score 663; DB 1; Length 198;
Best Local Similarity 61.6%; Pred. No. 2e-33;
Matches 117; Conservative 36; Mismatches 37; Indels 0; Gaps 0;

Qy 288 SVKLTSDFDNPRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347
Db 9 AVKLKTFDNDPRWIKRHKHMFDFLDINGNGKITLDEIVSKASDDICAKLGATPEQTKRHQ 68
Qy 348 DAVEAFFGAGMKYGVETDPAYTEGWKKLATDELYAKNEPTLIRIWGALFDIVDKD 407
Db 69 DAVEAFFGCGMEYKEITAFQFDGKQLATSELUKWARNEPTLIRIWGAQVDFIDKD 128
Qy 408 ONGATILDEWKAYTKAAGIIOSSEDCETFRVCDIDESSGOLDVDVDMTRHGLGFYWTMDPA 467
Db 129 GSGTITLDEWKAYKIGSISQEDCEATFRHCDLDSGDLVDVDMTRHGLGFYWTLDPN 188
Qy 468 CEKLYGGAVP 477
Db 189 ADGLYNGVP 198

RESULT 7
ZIC2_HUMAN STANDARD; PRT; 532 AA.
ID ZIC2_HUMAN STANDARD; PRT; 532 AA.
AC O95409; Q9H309;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein ZIC2 (Zinc finger protein of the cerebellum 2).
GN ZIC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. AND VARIANT HPES POLY-ALA INSERTION.
RX MEDLINE=98442655; PubMed=9771712;
RA Brown S.A., Warburton D., Brown L.Y., Yu C.Y., Roeder E.R.,
RA Stengel-Rutkowski S., Hennekam R.C., Muenke M.;
RT "Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila
RT odd-paired."
RL Nat. Genet. 20:180-183(1998).
RN [2]
SEQUENCE FROM N.A.
RX TISSUE=Brain;
RA MEDLINE=20556339; PubMed=10984499;
RA Yang Y., Hwang C.K., Junn E., Lee G., Mouradian M.M.;
RT "ZIC2 and Sp3 repress Sp1-induced activation of the human D1A dopamine
RT receptor gene."
RL J. Biol. Chem. 275:38863-38869(2000).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: DEFECTS IN ZIC2 ARE A CAUSE OF HOLOPROSENCEPHALY TYPE 5

```

```

CC (HPE5). HPE5 IS A STRUCTURAL ANOMALY OF THE BRAIN.
CC -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF104902; AAC96325.1; -
DR EMBL; AF193855; AAG28409.1; -
DR HSSP; P08047; 1SP2.
DR TRANSFAC; T04237; -
DR Genew; HGNC:12873; ZIC2.
DR MIM; 603073; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; 2f-C2H2; 5.
DR PRODOM; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Disease mutation; Holoprosencephaly.
FT DOMAIN 20 23 POLY-HIS.
FT DOMAIN 25 33 POLY-ALA.
FT DOMAIN 89 97 POLY-ALA.
FT DOMAIN 226 230 POLY-ALA.
FT DOMAIN 231 239 POLY-HIS.
FT DOMAIN 300 415 ZINC FINGERS.
FT ZN_FING 300 327 C2H2-TYPE (ATYPICAL).
FT ZN_FING 333 357 C2H2-TYPE.
FT ZN_FING 363 387 C2H2-TYPE.
FT ZN_FING 393 415 C2H2-TYPE.
FT DOMAIN 456 470 POLY-ALA.
FT DOMAIN 490 508 POLY-GLY.
FT VARIANT 470 470 A -> AAAAAAAAAA (IN HPE5).
FT CONFLICT 124 128 RGFGE -> ARLPGT (IN REF. 1).
FT SEQUENCE 532 AA; 55006 MW; BA3E6455DAF97EAC CRC64;

Query Match 6.7%; Score 171.5; DB 1; Length 532;
Best Local Similarity 31.0%; Pred. No. 0.0017;
Matches 54; Conservative 24; Mismatches 73; Indels 23; Gaps 8;

QY 124 ELKGDIFKEDCNILGHKLEYNVSHNYI--MADK--OKNGIKANFKIRHNIEDGSVQ- 177
Db 366 EFGCDRRFANSDDRKHKMHVHTSDKPYLCKMKDKSYTHPSSLRKHMKVHSSPQSSSS 425
QY 178 --LADHYQONTPIGDGPVLLPNHYLSTQSALSKDPNEKRDHMLVLEFVTAGITHGMDE 235
Db 426 PAASSGYESTPPG---LVSPS---AEPOSSSNLSP-----AAAAAAAAAAAVSA 473
QY 236 LYKSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 288
Db 474 VHRGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 525

RESULT 8
SWP1_ENCCU STANDARD; PRT; 450 AA.
AC Q9X2V1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Spore wall protein 1 precursor.
GN SWP1.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Bohne W., Ferguson D.J.P., Kohler K., Gross U.;
RT "Molecular characterisation of a developmentally expressed spore wall
RL protein from the human microsporidian Encephalitozoon cuniculi.";
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SPORE WALL COMPONENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ133745; CAB39735.1; -
KW Sporulation; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 450 SPORE WALL PROTEIN 1.
FT DOMAIN 66 74 POLY-ARG.
FT DOMAIN 79 82 POLY-SER.
FT DOMAIN 303 339 THR-RICH.
FT DOMAIN 355 450 GLY/SER-RICH.
FT SEQUENCE 450 AA; 45873 MW; 5E7071A3E3A6DF60 CRC64;

Query Match 6.3%; Score 161.5; DB 1; Length 450;
Best Local Similarity 55.4%; Pred. No. 0.0057;
Matches 41; Conservative 1; Mismatches 5; Indels 27; Gaps 4;

QY 239 SGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 274
Db 373 SDGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 432
QY 275 SG---GSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 285
Db 433 SGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 446

RESULT 9
SER1_GALME STANDARD; PRT; 115 AA.
AC O96614;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Sericin-1 (silk gum protein 1) (Fragment).
GN SER1 OR SER-1.
OS Galleria mellonella (Wax moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Pyraloidea; Pyralidae; Galleriinae; Galleria.
OX NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Middle silk gland;
RA Zurevec M., Sehnael F., Scheller K., Kumaran A.K.;
RT "Silk gland specific cDNAs from Galleria mellonella L.";
Insect Biochem. Mol. Biol. 22:55-67(1992).
CC -!- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
CC SECTION OF SILK GLANDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF095241; AAC79078.1; -

```







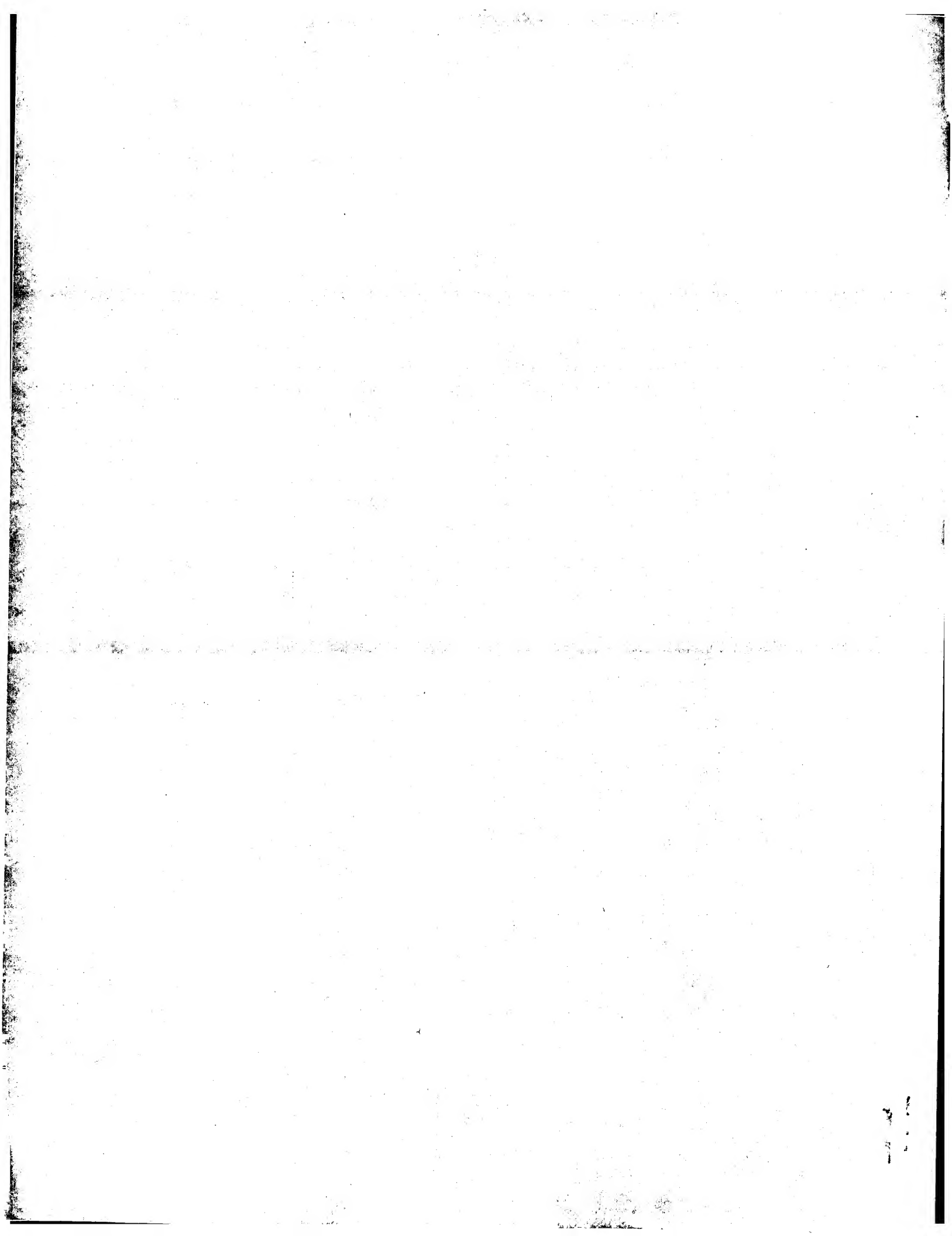




```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M34398; AAA39444.1; -.  
DR EMBL; U09189; AAA82152.1; -.  
DR PIR; A35628; A35628.  
DR HSSP; P02876; 9WGA.  
DR MGD; MGI:96816; Lor.  
KW Keratinization.  
SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;  
  
Query Match 5.9%; Score 151.5; DB 1; Length 481;  
Best Local Similarity 45.2%; Pred.No. 0.024; 33; Indels 11; Gaps 3;  
Matches 38; Conservative 2; Mismatches 3;  
  
QY 228 GITHGMDELYKSGGS-SGGQSGGS---SGQSGSGS-----SGQSGSGSGSGGQSG 276  
DB 395 GSGGGCGGYSGGGCGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSS 454  
  
QY 277 GSGSGGQSLRSVKLTSTDFDNPW 300  
DB 455 GGGSGGKGVPCVCHQTOQKQAPTW 478  
  
RESULT 14  
CORA_MEDSA  
ID CORA_MEDSA STANDARD; PRT; 204 AA.  
AC Q07202;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Cold and drought-regulated protein CORA.  
OS CORA.  
GN Medicago sativa (Alfalfa).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
OX NCBI_TaxID=3879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Apica;  
RX MEDLINE=94143496; PubMed=8310076;  
RA Laberge S., Castonguay Y., Vezina L.-P.;  
RT "New cold- and drought-regulated gene from Medicago sativa.";  
RL Plant Physiol. 101:1411-1412(1993).  
CC -!- FUNCTION: MAY BE INVOLVED IN RESISTANCE OF THE PLANT TO  
ENVIRONMENTAL STRESS.  
CC -!- INDUCTION: BY COLD, ABSCISIC ACID (ABA) AND DROUGHT STRESS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; L03708; AAA99833.1; -.  
DR MULTIGENE family; Repeat.  
KW Multigene family; Repeat.  
FT DOMAIN 54 176 7 X 6 AA REPEATS OF Y-N-H-G-G-G.  
FT REPEAT 54 59 1-1.  
FT REPEAT 65 70 1-2.  
FT REPEAT 71 76 1-3.  
FT REPEAT 78 83 1-4.  
FT REPEAT 85 90 1-5.  
FT REPEAT 164 169 1-6.  
FT REPEAT 171 176 1-7.  
FT DOMAIN 98 192 11 X 3 AA REPEATS OF H-G-G.  
FT REPEAT 98 100 2-1.  
FT REPEAT 101 103 2-2.  
FT REPEAT 112 114 2-3.  
FT REPEAT 115 117 2-4.  
FT REPEAT 126 128 2-5.
```

```
FT REPEAT 129 131 2-6.  
FT REPEAT 178 180 2-7.  
FT REPEAT 181 183 2-8.  
FT REPEAT 184 186 2-9.  
FT REPEAT 187 189 2-10.  
FT REPEAT 190 192 2-11.  
SQ SEQUENCE 204 AA; 19599 MW; 8A2C082359FCC17F CRC64;  
  
Query Match 5.8%; Score 149.5; DB 1; Length 204;  
Best Local Similarity 35.4%; Pred.No. 0.012;  
Matches 35; Conservative 11; Mismatches 44; Indels 9; Gaps 3;  
  
QY 228 GITHGMDELYKSGGS---GSGQSGSGSGSGSGSGSGS-GSGQSGSGSGSGSGSGSGGQ 283  
DB 70 GYNHGGGGYHNGGGYHNGGGGGHGGHGGGGYHNGGGGGYHNGGGGGYHNGGGGGHGGH 129  
  
QY 284 SGLRSVKL-----TSDFDNPWTRIGRHKHMFNFDVNHNG 317  
DB 130 GGAESVAVOTEKTEKNEVDKARYGGGSNYNDGRGYNHGG 168  
  
RESULT 15  
CANS_MOUSE  
ID CANS_MOUSE STANDARD; PRT; 269 AA.  
AC O88456; Q8VEK4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)  
DE (Calcium-activated neutral proteinase) (CANP).  
GN CAPNS1 OR CAPN4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20285452; PubMed=10825211;  
RA Arthur J.S.C., Elce J.S., Hegadorn C., Williams K., Greer P.A.;  
RT "Disruption of the murine calpain small subunit gene, Capn4; calpain  
RT is essential for embryonic development but not for cell growth and  
RT division.";  
RL Mol. Cell. Biol. 20:4474-4481(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which  
CC catalyze limited proteolysis of substrates involved in  
CC cytoskeletal remodeling and signal transduction.  
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small  
CC (regulatory) subunit.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma  
CC membrane upon Ca++ binding (By similarity).  
CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AF058298; AAC97194.1; -.  
DR EMBL; BC018352; AAH18352.1; -.  
DR HSSP; P04632; 1KFU.  
DR MGD; MGI:88266; Capns1.  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00036; ehand; 3.  
DR ProDom; PD000012; EF-hand; 1.  
DR PROSITE; PS00018; EF_HAND; 2.  
KW Calcium-binding; Repeat.
```





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:00:59 ; Search time 47.0664 Seconds  
(without alignments)  
2088.208 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGELFTGVVPIVLVDG.....LGFWTMDPCKLYGGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1256	49.1	238	5 Q93125	Q93125 aequorea vi
2	1254	49.0	238	5 Q27903	Q27903 unidentified
3	1231	48.1	758	4 Q96JQ1	Q96JQ1 homo sapien
4	1220	47.7	238	5 Q17105	Q17105 aequorea vi
5	1205	47.1	238	5 Q17106	Q17106 aequorea vi
6	1100	43.0	238	5 Q8WTC6	Q8WTC6 aequorea ma
7	1096	42.8	238	5 Q8WPC5	Q8WPC5 aequorea ma
8	1093	42.7	238	5 Q8WTC4	Q8WTC4 aequorea ma
9	1091	42.6	238	5 Q8WTD0	Q8WTD0 aequorea ma
10	1090	42.6	238	5 Q8WTC9	Q8WTC9 aequorea ma
11	1090	42.6	238	5 Q8WTC8	Q8WTC8 aequorea ma
12	1087	42.5	238	5 Q8WTC7	Q8WTC7 aequorea ma
13	1086	42.4	238	5 Q8WTC5	Q8WTC5 aequorea ma
14	915	35.7	195	5 Q8WQY8	Q8WQY8 aequorea pa
15	899	35.1	195	5 Q8WQY7	Q8WQY7 aequorea ma
16	705	27.5	195	5 Q8T6Z0	Q8T6Z0 obelia geni

17	253.5	9.9	225	5 Q95UA7	Q95UA7 montastraea
18	246	9.6	225	5 Q963F5	Q963F5 montastraea
19	240	9.4	236	5 Q8T6U0	Q8T6U0 dendronephth
20	233.5	9.1	266	5 Q9U6Y3	Q9U6Y3 clavularia
21	231.5	9.0	225	5 Q8T5F1	Q8T5F1 montastraea
22	210	8.2	227	5 Q962P9	Q962P9 montastraea
23	208	8.1	229	5 Q9U6Y6	Q9U6Y6 anemonia ma
24	207.5	8.1	234	5 Q8T5F2	Q8T5F2 montastraea
25	205.5	8.1	235	5 Q8T5F0	Q8T5F0 scolymia cu
26	205.5	8.0	232	5 Q9GP15	Q9GP15 anemonia su
27	204.5	8.0	225	5 Q9U6Y8	Q9U6Y8 discosoma s
28	204	8.0	227	5 Q95VT0	Q95VT0 montastraea
29	203.5	7.9	238	5 Q9BLY9	Q9BLY9 renilla mue
30	201.5	7.9	232	5 Q9GZ28	Q9GZ28 anemonia su
31	201	7.9	221	5 Q95P04	Q95P04 gonlopora t
32	193.5	7.8	225	5 Q8T6T9	Q8T6T9 heteractis
33	194	7.6	232	5 Q9U6Y7	Q9U6Y7 discosoma s
34	192	7.5	230	5 Q9GTJ7	Q9GTJ7 discosoma s
35	187.5	7.3	227	5 Q95W85	Q95W85 heteractis
36	185.5	7.3	228	5 Q9GP16	Q9GP16 anemonia su
37	182.5	7.1	238	5 Q9BLZ0	Q9BLZ0 ptilosarcus
38	181.5	7.1	233	5 Q963I9	Q963I9 renilla ren
39	179.5	7.0	227	5 Q95W86	Q95W86 condylactis
40	179.5	7.0	231	5 Q9U6Y5	Q9U6Y5 zoanthus sp
41	175	6.8	229	5 Q8T5E7	Q8T5E7 condylactis
42	173.5	6.8	227	5 Q95W11	Q95W11 condylactis
43	170	6.6	302	10 Q9SL09	Q9SL09 arabidopsis
44	170	6.6	738	5 Q02402	Q02402 pinctada fu
45	169	6.6	125	5 Q964C1	Q964C1 encephalito

#### ALIGNMENTS

RESULT 1

Q93125 ID Q93125 PRELIMINARY; PRT; 238 AA.  
AC Q93125;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Green fluorescent protein mutant 3.  
GN GFP.  
OS Aequorea victoria (Jellyfish).  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=6100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96305137; PubMed=8707053;  
RA Cornack B.P., Valdivia R.H., Falkow S.;  
RT "FACS-optimized mutants of the green fluorescent protein (GFP).";  
RL Gene 173:33-38(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Cornack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,  
Brown A.J.P.;  
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene  
expression in Candida albicans.";  
DR Microbiology 0:0-0(1996).  
DR EMBL; U73901; AAB18957.1; -.  
DR HSSP; P42212; 1BPP.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFLUORESCENT.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 49.1%; Score 1256; DB 5; Length 238;

Best Local Similarity 98.3%; Pred. No. 6e-82; Mismatches 3; Indels 0; Gaps 0;

Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGELFTGVVPIVLVDGDNVNGHKFSVSGEGEGDATYKLTCLKFICTTGTGLPVPWPTL 60

```
Db 1 MSKGEELFTGVVPLVLDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTFEGYGVQCFARYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHHVLTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHHVLTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 2
Q27903 ID Q27903 PRELIMINARY; PRT; 238 AA.
AC Q27903;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9729982; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
  plant Mol. Biol. 33:989-999(1997).
RL EMBL; X96418; CAA65278.1; -.
DR HSP; P42212; IGFL.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26884 MW; CA932D47262AF2D3 CRC64;

Query Match 49.0%; Score 1254; DB 5; Length 238;
Best Local Similarity 97.9%; Pred. NO. 8.3e-82;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
Db 1 MGKGEELFTGVVPLVLDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTFEGYGVQCFSRYPDHMKRHRDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHHVLTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHHVLTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 3
Q96J01 ID Q96J01 PRELIMINARY; PRT; 758 AA.
AC Q96J01;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Raichu404X.
GN RAICHU404X.
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RX MEDLINE=21322811; PubMed=11429608;
RA Mochizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
  Miyawaki A., Watsuda M.;
RT "Spacio-temporal Images of Growth Factor-induced Activation of Ras and
  Rapi.1";
RL Nature 411:1065-1068(2001).
DR EMBL; AB051846; BAB61868.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_transfmg.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; ras; 1.
DR Pfam; PF02196; RBD; 1.
DR ProDom; PD013756; Green_fl_protein; 2.
DR TIGRFAMS; TIGR00231; small_gtp; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.
SQ SEQUENCE 758 AA; 85015 MW; 8612408F607CFD49 CRC64;

Query Match 48.1%; Score 1231; DB 4; Length 758;
Best Local Similarity 94.3%; Pred. NO. 1.9e-79;
Matches 233; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 60
Db 499 VSKGEELFTGVVPLVLDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVWPPTL 558
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 559 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEGDTLV 618
QY 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 619 NRLELKGIDFKEDGNILGHKLEYNINSHNYYITADKQKNGIKANFKIRHNIEDGSVOLAD 678
QY 181 HYQONTPIGDPVLLPDNHHVLTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 679 HYQONTPIGDPVLLPDNHHVLTQSALS KDPNEKRDMHVLLEFVTAAGITLGMDELGRSR 738
QY 241 GSGSGGQ 247
Db 739 KMSKDGK 745

RESULT 4
Q17105 ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR HSP; P42212; IGFL.
DR InterPro; IPR000786; Green_fl_protein.
```

```
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFLUORESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 47.7%; Score 1220; DB 5; Length 238;
Best Local Similarity 94.5%; Pred. No. 2.2e-79;
Matches 225; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSKGELEFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGKLTLEFVTAAGTTHGMDELYK 60
DB 1 MSKGELEFTGVVPIILVELDGDVNGQKFSVSGEGGDATYGKLTLEFVTAAGTTHGMDELYK 60
QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHNIEDGSGVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKMEYNSHNVYIMGDKPKNGIKVNFKIRHNIKDGSGVOLAD 180
QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTTHGMDELYK 238
DB 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTTHGMDELYK 238

RESULT 5
Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR HSSP; P42212; 1BPP.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS: PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 47.1%; Score 1205; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2.6e-78;
Matches 223; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGELEFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGKLTLEFVTAAGTTHGMDELYK 60
DB 1 MSKGELEFTGVVPIILVELDGDVNGQKFSVSGEGGDATYGKLTLEFVTAAGTTHGMDELYK 60
QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHNIEDGSGVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKMEYNSHNVYIMGDKPKNGIKVNFKIRHNIKDGSGVOLAD 180
QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTTHGMDELYK 238
DB 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTTHGMDELYK 238
```

```
RESULT 6
Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 6B8FD75E88926903 CRC64;

Query Match 43.0%; Score 1100; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 8.1e-71;
Matches 197; Conservative 20; Mismatches 21; Indels 0; Gaps 0;

QY 1 MSKGELEFTGVVPIILVELDGDVNGHKFSVSGEGGDATYGKLTLEFVTAAGTTHGMDELYK 60
DB 1 MSKGELEFTGVVPIILVELDGDVNGHGFVSGEGGDATYGKLTLEFVTAAGTTHGMDELYK 60
QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
DB 61 VTTLYSGIOCFARYPEHMKMDFKSAMPEGVYQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHNIEDGSGVOLAD 180
DB 121 NRIELKGMDFKEDGNILGHKLEYNNSHNVYIMPKANNGLKVNFKIRHNIEDGSGVOLAD 180
QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLFEVTAAGTTHGMDELYK 238
DB 181 HIQTNVPLGDPVLPINHYLSTQTAISKDRNTRDHRHVMVLEFFSACGTHGMDELYK 238

RESULT 7
Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02062.1; -.
DR EMBL; AY013821; AAK02059.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
```

```
SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;
Query Match 42.8%; Score 1096; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.6e-70;
Matches 196; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGNGHKFVSVEGEGDATYKGLTKFICTTGGKLPVWPPTL 60
Db 1 MSKGEELFTGIVPVLIELDGDVGHGKFSVRGEGGADYKLEIKFICTTGGKLPVWPPTL 60
QY 61 VTTLTYGVCFSRYPDHMKOHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 61 VTTSYGILCFARYPEHMKWNDFKFSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY 121 NRIELKGIDKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMPDKNANGLKYNFKIRHNIEGGVOLAD 180
QY 181 HYQONTPIGGPVLPPDNHVLSTQSALS KDPNEKRDHMLVLEFVTAAGITHGMDLYK 238
Db 181 HYQTNVPLGDPVLIPIINHLYLSQTALSKDRNETRDHMLVLEFFSACGHTHGMDELYK 238

RESULT 8
Q8WTC4
ID Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHG24;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL3918.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 42.7%; Score 1093; DB 5; Length 238;
Best Local Similarity 83.6%; Pred. No. 2.6e-70;
Matches 199; Conservative 16; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGNGHKFVSVEGEGDATYKGLTKFICTTGGKLPVWPPTL 60
Db 1 MSKGEELFTGVVPIVLVDGNGHKFVSVRGEGGADYKLEIKFICTTGGKLPVWPPTL 60
QY 61 VTTLTYGVCFSRYPDHMKOHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 61 VTTLGYILCFARYPEHMKWNDFKFSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY 121 NRIELKGIDKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMPDKNANGLKYNFKIRHNIEGGVOLAD 180
QY 181 HYQONTPIGGPVLPPDNHVLSTQSALS KDPNEKRDHMLVLEFVTAAGITHGMDLYK 238
Db 181 HYQTNVPLGDPVLIPIINHLYLSQTALSKDRNETRDHMLVLEFFSACGHTHGMDELYK 238

RESULT 9
Q8WTD0
ID Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL3913.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;
```

```
Q8WTD0
ID Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL3912.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 42.6%; Score 1091; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 3.6e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGNGHKFVSVEGEGDATYKGLTKFICTTGGKLPVWPPTL 60
Db 1 MSKGEELFTGIVPVLIELDGDVGHGKFSVRGEGGADYKLEIKFICTTGGKLPVWPPTL 60
QY 61 VTTLTYGVCFSRYPDHMKOHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 61 VTTLGYILCFARYPEHMKWNDFKFSAMPEGYIQERTIFFQDDGKYKTRGEVKFEGDTLV 120
QY 121 NRIELKGIDKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMPDKNANGLKYNFKIRHNIEGGVOLAD 180
QY 181 HYQONTPIGGPVLPPDNHVLSTQSALS KDPNEKRDHMLVLEFVTAAGITHGMDLYK 238
Db 181 HYQTNVPLGDPVLIPIINHLYLSQTALSKDRNETRDHMLVLEFFSACGHTHGMDELYK 238

RESULT 10
Q8WTC9
ID Q8WTC9 PRELIMINARY; PRT; 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL3913.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;
```



```

Query Match      42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPLVIELDGDVGHGKFSVRGEGGDADYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VTTLTGVCQFSRYPDHMKQHDFFKFSAMPEGYQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLTGVCQFSRYPDHMKQHDFFKFSAMPEGYQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 NRLEKGDIDFKEDGNILGHKLEYNHSHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRLEKGMDFKEDGNILGHKLEYNHSHNYIMADKQKNGIKANFKIRHNIEGGVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLVLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLDGPVLLIPINHYLSQTALSADKPNKRDHMLVLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 11
Q8WTC8 PRELIMINARY; PRT; 238 AA.
AC Q8WTC8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPLVIELDGDVGHGKFSVRGEGGDADYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VTTLTGVCQFSRYPDHMKQHDFFKFSAMPEGYQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLTGVCQFSRYPDHMKQHDFFKFSAMPEGYQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 NRLEKGDIDFKEDGNILGHKLEYNHSHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRLEKGMDFKEDGNILGHKLEYNHSHNYIMADKQKNGIKANFKIRHNIEGGVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLVLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLDGPVLLIPINHYLSQTALSADKPNKRDHMLVLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 12
Q8WTC7 PRELIMINARY; PRT; 238 AA.
AC Q8WTC7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGGDATYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPLVIELDGDVGHGKFSVRGEGGDADYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VTTLTGVCQFSRYPDHMKQHDFFKFSAMPEGYQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLTGVCQFSRYPDHMKQHDFFKFSAMPEGYQVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 NRLEKGDIDFKEDGNILGHKLEYNHSHNYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRLEKGMDFKEDGNILGHKLEYNHSHNYIMADKQKNGIKANFKIRHNIEGGVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLVLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQTNVPLDGPVLLIPINHYLSQTALSADKPNKRDHMLVLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 13
Q8WTC5 PRELIMINARY; PRT; 238 AA.
AC Q8WTC5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Orange fluorescent protein.
GN Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OPPM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435432; AAL33917.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;

Query Match      42.4%; Score 1086; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 8.1e-70;

```

Matches 198; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

```

QY 1 MSGGELFTGVVPIILVELDGVNGHKSFSVSGEGDGYATYGLTKLFICTTGKLPVPWPTL 60
DB 1 MSGGELFTGVVPIILVELDGVNGHKSFSVSGEGDGYATYGLTKLFICTTGKLPVPWPTL 60
QY 61 VTTLYGVOCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNKYKTRAEVKEGFTLV 120
DB 61 VTTLYGVOCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNKYKTRAEVKEGFTLV 120
QY 121 NRIELKGTDFEDNGNGLHKEYNWNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGTDFEDNGNGLHKEYNWNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQNTPTGDPVLLPNHLYLSQSALSQPNKRDHMLVLEFVTAAGITHGMDELYK 238
DB 181 HYQNTPTGDPVLLPNHLYLSQSALSQPNKRDHMLVLEFVTAAGITHGMDELYK 238

```

## RESULT 14

```

Q8WQY8
ID Q8WQY8 PRELIMINARY; PRT; 195 AA.
AC Q8WQY8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Apoaequorin.
GN Aequorea.
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=148610;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AEQXMX;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,
RA Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; EF-hand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 3.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_3.
SQ SEQUENCE 195 AA; 22248 MW; 73EB72E7E0C140BE CRC64;

```

Query Match 35.7%; Score 915; DB 5; Length 195;  
 Best Local Similarity 85.3%; Pred. No. 9.7e-56;  
 Matches 162; Conservative 18; Mismatches 10; Indels 0; Gaps 0;

```

QY 288 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 347
DB 6 AVKLEPDPENPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 65
QY 348 DAVEAFFGGAGMKYGVETDWPAYTEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
DB 66 EAVEAFFGGAGMKYGVETDWPAYTEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETPRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467
DB 126 QNGAITLDEWKAYTKAAGIIQSSDCEETPRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467
QY 468 CEKLYGGAVP 477
DB 186 CEKLYGGAVP 195

```

## RESULT 15

```

Q8WQY7
ID Q8WQY7 PRELIMINARY; PRT; 195 AA.
AC Q8WQY7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

```

```

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Apoaequorin.
GN Aequorin.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AEQXMX;
RA Luo W.X., Zhang J., Yang H.J., Li S.W., Xie X.Y., Qin Y.X., Pang S.Q.,
RA Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013823; AAK02060.1; -.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; EF-hand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_3.
SQ SEQUENCE 195 AA; 22266 MW; A459391B4B8003BF CRC64;

```

Query Match 35.1%; Score 899; DB 5; Length 195;  
 Best Local Similarity 83.7%; Pred. No. 1.4e-56;  
 Matches 159; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

```

QY 288 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 347
DB 6 AVKLEPDPENPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 65
QY 348 DAVEAFFGGAGMKYGVETDWPAYTEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
DB 66 DAVEAFFGGAGMKYGVETDWPAYTEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETPRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467
DB 126 QNGAITLDEWKAYTKAAGIIQSSDCEETPRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467
QY 468 CEKLYGGAVP 477
DB 186 CEKLYGGAVP 195

```

Search completed: March 13, 2003, 15:05:05  
 Job time : 50.0664 secs